

SEARCH REQUEST FORM

Scientific and Technical Information Center

Requester's Full Name: _____ Examiner #: _____ Date: _____
 Art Unit: _____ Phone Number 30 _____ Serial Number: _____
 Mail Box and Bldg/Room Location: _____ Results Format Preferred (circle): PAPER DISK E-MAIL

If more than one search is submitted, please prioritize searches in order of need.

Please provide a detailed statement of the search topic, and describe as specifically as possible the subject matter to be searched. Include the elected species or structures, keywords, synonyms, acronyms, and registry numbers, and combine with the concept or utility of the invention. Define any terms that may have a special meaning. Give examples of relevant citations, authors, etc, if known. Please attach a copy of the cover sheet, pertinent claims, and abstract.

Title of Invention: _____

Inventors (please provide full names): _____

Earliest Priority Filing Date: _____

**For Sequence Searches Only* Please include all pertinent information (parent, child, divisional, or issued patent numbers) along with the appropriate serial number.*

STAFF USE ONLY

Contact: _____

Searcher: Shoppard

Searcher Phone #: rel. 308-4499

Searcher Location: _____

Date Searcher Picked Up: _____

Date Completed: 9/3/00

Searcher Prep & Review Time: _____

Clerical Prep Time: _____

Online Time: _____

Type of Search

NA Sequence (#) _____

AA Sequence (#) _____

Structure (#) _____

Bibliographic _____

Litigation _____

Fulltext _____

Patent Family _____

Other _____

Vendors and cost where applicable

STN _____

Dialog _____

Questel/Orbit _____

Dr.Link _____

Lexis/Nexis _____

Sequence Systems _____

WWW/Internet _____

Other (specify) _____

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 1, 2000, 23:12:56 ; Search time 2111.72 Seconds
(without alignments)
2743.978 Million cell updates/sec

Title: US-09-464-528-6
Perfect score: 1314
Sequence: 1 tctagatcaaacacacatcc.....ttttgaagtataaccatgg 1314

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 5247842 seqs, 2204914090 residues

Total number of hits satisfying chosen parameters: 10495684

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST: *
1: em_est1: *
2: em_est2: *
3: em_est3: *
4: em_est4: *
5: em_est5: *
6: em_est6: *
7: em_est7: *
8: em_est8: *
9: em_est9: *
10: em_est10: *
11: em_est11: *
12: em_est12: *
13: em_est13: *
14: em_est14: *
15: em_est15: *
16: em_est16: *
17: em_est17: *
18: em_est18: *
19: em_est19: *
20: gb_est1: *
21: gb_est2: *
22: gb_est3: *
23: gb_est4: *
24: gb_est5: *
25: gb_est6: *
26: gb_est7: *
27: gb_est8: *
28: gb_est9: *
29: gb_est10: *
30: gb_est11: *
31: gb_est12: *
32: gb_est13: *
33: gb_est14: *
34: gb_est15: *
35: gb_est16: *
36: gb_est17: *
37: gb_est18: *
38: gb_est19: *
39: gb_est20: *
40: gb_est21: *
41: gb_est22: *
42: gb_est23: *
43: gb_est24: *

44: gb_est25: *
45: gb_est26: *
46: gb_est27: *
47: gb_est28: *
48: gb_est29: *
49: gb_est30: *
50: gb_est31: *
51: gb_est32: *
52: em_est20: *
53: em_est21: *
54: em_est22: *
55: em_est23: *
56: em_est24: *
57: em_est25: *
58: em_est26: *
59: gb_est33: *
60: gb_est34: *
61: gb_est35: *
62: gb_est36: *
63: gb_est37: *
64: gb_est38: *
65: em_est27: *
66: em_est28: *
67: em_est29: *
68: em_est30: *
69: gb_est39: *
70: gb_est40: *
71: gb_est41: *
72: gb_est42: *
73: gb_est43: *
74: gb_est44: *
75: em_est31: *
76: em_est32: *
77: em_est33: *
78: em_est34: *
79: gb_est45: *
80: gb_est46: *
81: gb_est47: *
82: em_est35: *
83: em_est36: *
84: em_est37: *
85: gb_est48: *
86: gb_est49: *
87: gb_est50: *
88: gb_est51: *
89: gb_est52: *
90: gb_est53: *
91: gb_est54: *
92: gb_est55: *
93: gb_est56: *
94: gb_est57: *
95: gb_est58: *
96: gb_est59: *
97: em_gss1: *
98: em_gss2: *
99: em_gss3: *
100: em_gss4: *
101: gb_gss5: *
102: gb_gss6: *
103: gb_gss7: *
104: gb_gss8: *
105: gb_gss9: *
106: em_gss5: *
107: em_gss6: *
108: em_gss7: *
109: em_gss8: *
110: em_gss9: *
111: em_gss10: *
112: em_gss11: *
113: gb_gss10: *
114: gb_gss11: *
115: em_gss12: *
116: gb_gss12: *

```

source
1. .843
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="RPIC-98"
/clone="BACR26H19"
/note="end : TET3"
BASE COUNT 131 a 102 c 259 g 141 t 210 others
ORIGIN

Query Match 6.4%; Score 84.6; DB 122; Length 843;
Best Local Similarity 39.2%; Pred. No. 11e-05;
Matches 121; Conservative 65; Mismatches 123; Indels 0; Gaps 0;

QY 1 tctagatcaaacatcacatccaaacataacatgatatctcttaccatacatactaatta 60
Db 831 TAWAWATAAATATATATWTWTTTTATAATTTDWTTWDAATAAMWTATATWATTTT 772
QY 61 ttttggggttaataataataatatttttaagaataataataaagaataataaagaatttt 120
Db 771 TTTTAAWTWTWTWTWTATTTAAATTTWTATATATTTATATATTTTAAATTTTWTAT 712
QY 121 taataaaatgcataaaattatattcattcattgattttccatacatattgatttgcataaa 180
Db 711 ATTAWATAATATATAAWATAAATTAATTAATTTTAAWAWWWTATTTTATTTTATTTTATAT 652
QY 181 atatatattttttaaattctttaaataaagtgtgcagacacattattagacatagctctgtt 240
Db 651 TTWTWTWTATAAATTAATTTAAATTAATTAATTAATTAATTAATTTTWTRTAAAAAAMWAAAA 592
QY 241 ctgtttacaagaagcattcatttataacattataaataataataataacataacagtagaaa 300
Db 591 AATWATATAAAAAAAMWATAAHTTTTATMAAATTTWAAATAATAAATAAATAAATAAAMWAA 532
QY 301 tctctctgt 309
Db 531 ATWATWAT 523

RESULT 3
CNS016LI/c
LOCUS
DEFINITION
Drosophila melanogaster genome survey sequence T7 end of BAC
BACN16D22 of DrosBAC library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION
AL106896
VERSION
GI:5624374
KEYWORDS
GSS.
SOURCE
fruit fly.
ORGANISM
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 1101)
Genoscope.
Direct Submission
Submitted (23-JUL-1999); Genoscope - Centre National de Sequencage;
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
collaboration with the European Drosophila Genome Project (EDGP).
http://www.edgp.ebi.ac.uk - This Drosophila melanogaster BAC
library (Dros BAC) was made by Alain Billaut at CEPH (Centre
d'Etude du Polymorphisme Humain) with funding provided by a MRC
project grant. The DNA was prepared from embryos by Alain Bucheton
and Genevieve Payan. It has been constructed in the vector
pbeloBAC11.
FEATURES
Location/Qualifiers
1. .1101
/organism="Drosophila melanogaster"
/plasmid="pbeloBAC11"
/db_xref="taxon:7227"
/clone_lib="DrosBAC"

```


Matches	115;	Conservative	57;	Mismatches	94;	Indels	3;	Gaps	1;
---------	------	--------------	-----	------------	-----	--------	----	------	----

QY	48	atcatacgaattttgggtgtaaaataaatcaattattttttaagatatcaataaataaaa	107
		: : : : : : : : : :	
Db	1197	ATTWAMTTATWTATTATATWTATAATWWAAWMTTAAAAAATAATAATATTAATAAAWA	1138
QY	108	ttaagaagtctttaaataaagtataataatataattcatgatgttttcacacatttg	167
		: : : : : : : : : :	
Db	1137	ATWTATGTAWTTTTTAAWAATATWTATWTATWTATTAATWMTTATTAATWMTTTTTTTTTTTMMWTTT	1078
QY	168	atttgtaataaatacatttttttttttttcttcaaaaaatgtgcgaagcacct--tat	224
		: : : : : : : : : : :	
Db	1077	TWTTTTTWTAATWTWTTTTTTTTTWTWAAWMTWTWTWHAATWTATTTTATATTATTAATAAW	1018
QY	225	tagacatagcttggtctgttttacaaaagcattcatcatttaacattaaaaaatattt	284
		: : : : : : : : : : : :	
Db	1017	TTTWTATWTWTATWTATWTATWTAAGAAAATAATTTTTTWMTTTTAAAAATTAAARAWAAWATTA	958
QY	285	aactaacagtagaactctctctgtgagt	313
		: : : : : : : : : : :	
Db	957	AATTWAAAAAAWAWTTTTTTTTRTTTWWAAAT	929

RESULT	7
CNSOODKY/c	
LOCUS	
DEFINITION	CNSO0DKY 928 bp DNA GSS 04-JUN-1999
	Drosophila melanogaster genome survey sequence T7 end of BAC #
	BACR27A24 of RPCI-98 library from Drosophila melanogaster (fruit
	fly), genomic survey sequence.
ACCESSION	AL071865
VERSION	AL071865.1 GI:4948170
KEYWORDS	GSS.
SOURCE	fruit fly.
ORGANISM	Drosophila melanogaster
	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
	Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE	1 (bases 1 to 928)
AUTHORS	Genoscope.
TITLE	Direct Submission
JOURNAL	Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage ;
	BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
	- Web : www.genoscope.cns.fr)
COMMENT	determination of this BAC-end sequence was carried out as part of a
	collaboration with the Berkeley Drosophila Genome Project (BDGP).
	The BDGP is constructing a physical map of the Drosophila
	melanogaster genome using these BACs. For further information
	please see http://www.fruitfly.org The BDGP Drosophila
	melanogaster BAC library was prepared by Kazutoyo Osoegawa and
	Aaron Mammossier in Pieter de Jong's laboratory in the Department of
	Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
	NY. The library is named RPCI-98 and was constructed by partial
	EcoRI digestion of Drosophila DNA provided by the BDGP from the
	isogenic strain Y2; cn bw sp, the same strain used for the BDGP's
	PI and EST libraries. A more detailed description of the library
	and how to order individual BAC clones, the entire library, or
	filters for hybridization from the BACPAC Resource Center can be
	found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES	Location/Qualifiers
source	1..928
	/organism="Drosophila melanogaster"
	/db_xref="taxon:7227"
	/clone.lib="RPCI-98"
	/clone="BACR27A24"
	/note="end : T7"
BASE COUNT	262 a 70 c 321 t 191 others
ORIGIN	

Query Match	6.0%;	Score	78.8;	DB	122;	Length	928;
Best Local Similarity	29.7%;	Pred.	No. 0.00011;				
Matches	91;	Conservative	99;	Mismatches	116;	Indels	0;
						Gaps	0;

[illegible]

	Matches	129:	Conservative	49:	Mismatches	147:	Indels	0:	Gaps	0:
Qy	3	tagatcaaaac	ccatccaacatacaac	atgagatccctt	accacatcatactaatt	tatt	62			
		:	:	:	:	:				
Db	961	WAAAAAAATTTAATAW	WAAAAAAWAAAAAAW	ATAATATATAATAA	WWTAAATWTTAAATW	TWAT	902			
		:	:	:	:	:				
Qy	63	ttgggttaaatataatt	cattatttttaagata	tatttaaga	aaattcaaaagattttta	122				
		:	:	:	:	:				
Db	901	TWAAAAAATAATAWA	ATAWAAATTA	WAAAAAATAWATA	WAAAAAATWATA	TAATWATA	842			
		:	:	:	:	:				
Qy	123	aaaaaatgtaaaaaat	tattatttcagattttc	tacatttggttttgata	ataataaat	182				
		:	:	:	:	:				
Db	841	AAAAATWTTWAWWAAA	AAATWTTTTTAAATAT	TATATATWTTAAAT	WTTAAATWTTWAAAAA	AAAT	782			
		:	:	:	:	:				
Qy	183	atatatttttaattct	tataaaaaatgttgcag	acactattagacat	agctctgtct	242				
		:	:	:	:	:				
Db	781	AWAAWATAWTTWAWW	WNAWAAAAATAATAT	ATWNAWAAAAAAT	TWAWWNAATWTTW	722				
		:	:	:	:	:				
Qy	243	gtttacaaagcattca	tcatcttaatacacatt	taaaaaatttatactac	taacagtaga	302				
		:	:	:	:	:				
Db	721	WTAATAATAAAAAA	ATAWATAATWAAAA	AAAAAATAAAAAA	AAAAAATAAAAAA	662				
		:	:	:	:	:				
Qy	303	ttctgtgagtggtg	tggtgagtagg	327						
		:	:	:	:	:				
Db	661	AAGNNGGGGGGGGG	GGGGGGGGGGGGGG	GGGGGGGGGGGGGG	GGGGGGGGGGGGGG	637				
		:	:	:	:	:				

RESULT 14
 CNS00DKY
 LOCUS
 DEFINITION
 CNS00DKY 928 bp DNA GSS 04-JUN-1999
 Drosophila melanogaster genome survey sequence T7 end of BAC #
 BACR37A34 of RPCI-98 library from *Drosophila melanogaster* (fruit
 fly), genomic survey sequence.
 ACCESSION AL071865
 VERSION AL071865
 KEYWORDS GSS.
 SOURCE fruit fly.
 ORGANISM *Drosophila melanogaster*
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
 Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 1 (bases 1 to 928)
 Genoscope.
 Direct Submission
 Submitted (02-JUN-1999) Genoscope - Centre National de Sequencag
 BP 191 91006 EVRY cedex - FRANCE (E-mail : segref@genoscope.cns.fr)
 Web : www.genoscope.cns.fr)
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL

COMMENT

WEB : www.genoscope.cns.fr

Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osogawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's PI and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

```

found at location/Qualifiers
FEATURES
source
1. 928
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/Clone_lib="RPCI-98"
/Clone="BACR37A24"
/note="end : T7"
262 a 70 c 84 g 321 t 191
BASE COUNT
ORIGIN

```

[illegible]

RESULT 15
CNS00FUH/c
LOCUS
DEFINITION
 CNS00FUH 996 bp DNA GSS 03-JUN-1999
Drosophila melanogaster genome survey sequence TET3 end of BAC:
BACR31021 of RPC1-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCION
VERSION AL071063 1 GI:4951105
KEYWORDS
SOURCE fruit fly.
 Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
 Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
 1 (bases 1 to 996).
AUTHORS
TITLE Direct Submision
JOURNAL
 Submitted (02-JUN-1999), Genoscope - Centre National de Sequencage
 BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)
 - Web : www.genoscope.cns.fr)
COMMENT
 Determination of this BAC-end sequence was carried out as part of

COMMENT

Web : www.genoscope.cns.fr
Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the *Drosophila melanogaster* genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP *Drosophila melanogaster* BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPQJ-98 and was constructed by partial EcoRI digestion of *Drosophila* DNA provided by the BDGP from the isogenic strain y2; cn bw sp; the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

```

FEATURES
  source
    1. .996
      Location/Qualifiers
        db_xref="taxon:7227"
        clone_lib="PC1-98"
        clone="BACR31021"
        notes="end : TE33"
BASE COUNT      383 a      164 c      81 g      171 t      197 others
ORIGIN
  Query Match      5.7% ; Score 74.4 ; DB 122; Length
  Best Local Similarity 41.0% ; Pred. No. 0.00064;

```

```
Matches 109; Conservative 50; Mismatches 107; Indels 0; Gaps 0;
Qy 26 taacatggatctctccatccatccatcaatatttttgggttaaatattatcaattat 85
   | : | : | | | | | | | | | | | | | | | | | | | | | | | |
Db 863 TTAAATAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 804
   | : | : | | | | | | | | | | | | | | | | | | | | | | | |
Qy 86 ttttaagatataataaagaatttaaaagattttttaaaaaaagtataaaattatatta 145
   | : | : | | | | | | | | | | | | | | | | | | | | | | | |
Db 803 AAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 744
   | : | : | | | | | | | | | | | | | | | | | | | | | | | |
Qy 146 ttcatgatttttccatcacatttgatttgataataataatatttttttcttcttaaaa 205
   | : | : | | | | | | | | | | | | | | | | | | | | | | | |
Db 743 TTTTAAATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 684
   | : | : | | | | | | | | | | | | | | | | | | | | | | | |
Qy 206 aatgttgcaagacacttattagacatagctctgtctgtttcacaaaaagcatcattt 265
   | : | : | | | | | | | | | | | | | | | | | | | | | | | |
Db 683 TTTATTTAAATGATTTAAATTAATTAATTAATTAATTAATTAATTAATTA 624
   | : | : | | | | | | | | | | | | | | | | | | | | | | | |
Qy 266 aatcattaaaaaataatttaacta 291
   | : | : | | | | | | | | | | | | | | | | | | | | | | | |
Db 623 WTTAKAAATTTWKAATTTTGAATAATW 598
```

Search completed: September 2, 2000, 00:24:41
Job time: 4305 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 1, 2000, 23:40:36 : Search time 111.97 seconds
(without alignments)
1614.257 Million cell updates/sec

Title: US-09-464-528-6
Perfect score: 1314
Sequence: 1 tctagatcaaacacacaccc.....ttttgaagtataaccatgg 1314

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 243080 seqs, 68777915 residues

Total number of hits satisfying chosen parameters: 486160

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*

- 1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
- 2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
- 3: /cgn2_6/ptodata/1/ina/5C_COMB.seq:*
- 4: /cgn2_6/ptodata/1/ina/5D_COMB.seq:*
- 5: /cgn2_6/ptodata/1/ina/6_COMB.seq:*
- 6: /cgn2_6/ptodata/1/ina/PCITUS_COMB.seq:*
- 7: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	78	5.9	7218	1	US-08-232-463-14
2	56	4.3	19124	4	US-08-487-826B-13
3	54.2	4.1	1186	4	US-08-731-722-5
4	53.6	4.1	19124	4	US-08-487-826B-13
5	52.6	4.0	665	4	US-08-883-795A-36
6	52	4.0	8920	3	US-08-446-855A-1
7	51.2	3.9	3933	4	US-08-731-722-3
8	51.2	3.9	3933	4	US-08-731-722-3
9	51	3.9	665	4	US-08-883-795A-36
10	50.8	3.9	602	1	US-08-229-393-1
11	50.2	3.8	473	2	US-08-764-100-16
12	50.2	3.8	4970	2	US-08-764-100-14
13	50.2	3.8	4970	2	US-08-764-100-20
14	50.2	3.8	8920	3	US-08-446-855A-1
15	49.8	3.8	4467	2	US-08-565-907A-1
16	49.8	3.8	4467	2	US-08-910-551B-1
17	49.8	3.8	4467	4	US-08-909-425A-1
18	49.4	3.8	1611	7	5213972-6
19	49.2	3.7	319	1	US-07-593-657-14
20	49	3.7	4098	4	US-08-605-106-4
21	48.2	3.7	6768	2	US-08-107-755A-1
22	48.2	3.7	8457	1	US-07-991-867B-1
23	48.2	3.7	8457	4	US-08-544-332-1
24	47.8	3.6	12124	1	US-08-181-271A-36
25	47.8	3.6	12124	1	US-08-449-315-36
26	47.8	3.6	12124	1	US-08-444-803-36

27 47.8 3.6 12124 1 US-08-449-043-36 Sequence 36, Appl
28 47.8 3.6 12124 2 US-08-456-265A-36 Sequence 36, Appl
29 47.8 3.6 12124 2 US-08-455-416-36 Sequence 36, Appl
30 47.8 3.6 12124 2 US-08-455-244-36 Sequence 36, Appl
31 47.8 3.6 12124 2 US-08-454-876-36 Sequence 36, Appl
32 47.8 3.6 12124 3 US-08-457-364-36 Sequence 36, Appl
33 47.8 3.6 12124 3 US-08-456-262-36 Sequence 36, Appl
34 47.8 3.6 12124 3 US-08-456-240-36 Sequence 36, Appl
35 47.8 3.6 12124 3 US-08-455-736-36 Sequence 36, Appl
36 47.8 3.6 12124 4 US-08-971-217-36 Sequence 36, Appl
37 47.6 3.6 10385 1 US-08-245-809-3 Sequence 3, Appl
38 47.6 3.6 10396 1 US-08-245-809-5 Sequence 5, Appl
39 47.6 3.6 10798 1 US-08-107-748-2 Sequence 2, Appl
40 47.6 3.6 10798 6 PCT-US92-01385-2 Sequence 2, Appl
41 47.6 3.6 10965 1 US-08-107-748-4 Sequence 4, Appl
42 47.6 3.6 10965 6 PCT-US92-01385-4 Sequence 4, Appl
43 47.4 3.6 642 2 US-08-764-100-13 Sequence 13, Appl
44 47.4 3.6 643 2 US-08-764-100-7 Sequence 7, Appl
45 47.4 3.6 2993 2 US-08-764-100-2 Sequence 2, Appl

ALIGNMENTS

RESULT 1
US-08-232-463-14
Sequence 14, Application US/08232463
Patent No. 5670367
GENERAL INFORMATION:
APPLICANT: DORNER, F.
APPLICANT: SCHEIFLINGER, F.
APPLICANT: FALKNER, F. G.
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 1800 Diagonal Road, Suite 500
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA: US/08/232,463
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/935,313
FILING DATE:
APPLICATION NUMBER: EP 91 114 300.6
FILING DATE: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 30472/114 IMM
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)836-9300
TELEFAX: (703)683-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: pTZgpt-Fls
US-08-232-463-14


```
Query Match 5.9%; Score 78; DB 1; Length 7218;
Best Local Similarity 3.8%; Pred. No. 7.5e-07;
Matches 15; Conservative 241; Mismatches 136; Indels 0; Gaps 0;

Qy 554 ctttcgcgttaactacccctgccaccggtttccctataaattggaactcaatgct 613
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1071 yvvvyyvyyvyyvyyvyyvyyvyyvyyvyyvyyvyyvyyvyyvyyvyyvyy 1130

Qy 614 cccctctaaactcgatcgcttcagagtgagacacacacacacgcttcatactct 673
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1131 yvvvyyvyyvyyvyyvyyvyyvyyvyyvyyvyyvyyvyyvyyvyyvyyvyy 1190

Qy 674 ctgctctctctctctctctcctcctcaaggtactttctctcctcctcacaatccta 733
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1191 yvvvyyvyyvyyvyyvyyvyyvyyvyyvyyvyyvyyvyyvyyvyyvyyvyy 1250

Qy 734 gattccggtggtcaattcgatctgcaactctggttgccttgcctgctttctctc 793
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1251 yvvvyyvyyvyyvyyvyyvyyvyyvyyvyyvyyvyyvyyvyyvyyvyyvyy 1310

Qy 794 aactgggtccatcgatcgatcgaaactctactcttcttcttaatactcggaatac 853
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1311 yvvvyyvyyvyyvyyvyyvyyvyyvyyvyyvyyvyyvyyvyyvyyvyyvyy 1370

Qy 854 gcgttggaacttcagatctagtcgaaactcattcacaattgcttcttctttagctt 913
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1371 yvvvyyvyyvyyvyyvyyvyyvyyvyyvyyvyyvyyvyyvyyvyyvyyvyy 1430

Qy 914 atgagaaataaaactcatttttttttatttcaa 945
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1431 yvvvgtacaaattctctatctcttctaacta 1462

RESULT 2
US-08-487-826B-13
; Sequence 13, Application US/08487826B
; Patent No. 5993827
; GENERAL INFORMATION:
; APPLICANT: Sim, Kim L.
; APPLICANT: Chitnis, Chetan
; APPLICANT: Miller, Louis H.
; APPLICANT: Peterson, David S.
; APPLICANT: Su, Xin-zhaun
; APPLICANT: Wellem, Thomas E.
; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
; TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe Martens Olson & Bear
; STREET: 620 Newport Center Drive 16th Floor
; CITY: Newport Beach
; STATE: California
; COUNTRY: US
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/487.826B
; FILING DATE: 10-SEP-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Israelsen, Ned
; REGISTRATION NUMBER: 29,655
; REFERENCE/DOCKET NUMBER: NIH121.001CP1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 235-8550
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
```

```
LENGTH: 19124 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-487-826B-13

Query Match 4.3%; Score 56; DB 4; Length 19124;
Best Local Similarity 47.7%; Pred. No. 0.012;
Matches 164; Conservative 0; Mismatches 180; Indels 0; Gaps 0;

Qy 57 attatttgggttaataataataataataataataataataataataataataata 116
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1264 ATTATAATATGTAATATTATAATAATATATTGTATACATACACAGCAATAAGAAC 1323

Qy 117 tttttaaataatgtataataataataataataataataataataataataataata 176
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1324 TATACAATCTGCTATCTAATAGTATATATATATATATATCTTTTATTATTTGTTCT 1383

Qy 177 ataaataatatttttttaatttttaataataataataataataataataataata 236
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1384 CTTTCTTTTCTTTTAAATAATAATAATAATAATAATAATAATAATAATAATAATA 1443

Qy 237 tggctgtttcacaaagcattcattcattcattcattcattcattcattcattcatt 296
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1444 TAGTATTTTATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 1503

Qy 297 agaattcttctgtgagtggtggtggtggtggtggtggtggtggtggtggtggtggt 356
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1504 ATATTAATATTAGTAACCCACCTAGATAAATTAGAGAGAAACGTAGACATACCA 1563

Qy 357 gtcagaaacagaaagcaataataataataataataataataataataataata 400
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1564 ATTAGAACAAAAGAAATATTACAAAAAATAATAATAATAATAATAATAATA 1607

RESULT 3
US-08-731-722-5/c
; Sequence 5, Application US/08731722
; Patent No. 5961971
; GENERAL INFORMATION:
; APPLICANT: Martin, Frank N.
; TITLE OF INVENTION: Biocontrol of Fungal Soilborne Pathogens
; TITLE OF INVENTION: by Pythium oligandrum
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Saliwanchik & Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: US
; ZIP: 32608-6669
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/731.722
; FILING DATE:
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Whitlock, Ted W.
; REGISTRATION NUMBER: 36,965
; REFERENCE/DOCKET NUMBER: UF-161
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 352-375-8100
; TELEFAX: 352-372-5800
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
```

LENGTH: 1186 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
INDIVIDUAL ISOLATE: 23-5
US-08-731-722-5

Query Match 4.1%; Score 54.2; DB 4; Length 1186;
Best Local Similarity 59.3%; Pred. No. 0.018;
Matches 112; Conservative 0; Mismatches 73; Indels 4; Gaps 1;

Qy 52 tactaattatttgggttaataatcatcattatttttaagataattaataagaataaa 111
Db 595 TATTGACGAATTTCTTAACACTACTTAAGATTTTAGGACTACTGTATAATCCAG 536
Qy 112 aagattttttaaataatgtataaaattatattattcatgattttttcatatcatttgattt 171
Db 535 ATCATATTATAAATAATAAAGTTT---TTCAATATTATAATAAATAATATATTC 480
Qy 172 tgataataaatatatttttttatttttttttaaaatgttgcaagacacttattagacat 231
Db 479 TTATTAGAAGTATTTTCATCTTTAATTTT---TTAAAGTTATATATCTTTAAAGATAT 420
Qy 232 agtcctgtt 240
Db 419 AATTAAT 411

RESULT 4
US-08-487-826B-13/C
Sequence 13, Application US/08487826B
Patent No. 5993827
GENERAL INFORMATION:
APPLICANT: Sim, Kim L.
APPLICANT: Chitnis, Chetan
APPLICANT: Miller, Louis H.
APPLICANT: Peterson, David S.
APPLICANT: Su, Xin-zhaun
APPLICANT: Wellem, Thomas E.
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESS:
ADDRESSEE: Knobbe Martens Olson & Bear
STREET: 620 Newport Center Drive 16th Floor
CITY: Newport Beach
STATE: California
COUNTRY: US
ZIP: 92660

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/487,826B
FILING DATE: 10-SEP-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Israel, Ned
REGISTRATION NUMBER: 29,655
REFERENCE/DOCKET NUMBER: NIH121.001CP1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 19124 base pairs
TYPE: nucleic acid
STRANDEDNESS: single

TOPOLOGY: linear
MOLECULE TYPE: CDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-487-826B-13

Query Match 4.1%; Score 53.6; DB 4; Length 19124;
Best Local Similarity 45.9%; Pred. No. 0.033;
Matches 221; Conservative 0; Mismatches 259; Indels 2; Gaps 1;

Qy 818 tgaactactactcttcttcttaataatctgcggaatacgcgttggacttccagatcgtcgc 877
Db 15940 TTAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 15881
Qy 878 aaatcattcattgacctcttcttcttcttcttcttcttcttcttcttcttcttctt 937
Db 15880 AATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 15821
Qy 938 tattcaaaaataaaccttggccttgcgtgactgagatggggttgcgtgactacagaat 997
Db 15820 TATTTT---AATAAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 15763
Qy 998 ttgagcgaatttgcgaattgactgttgcgtgactgttgcgtgactgttgcgtgactgt 1057
Db 15762 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 15703
Qy 1058 catactccttagcttcaatttttcttcttcttcttcttcttcttcttcttcttctt 1117
Db 15702 TTTATGATATATATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 15643
Qy 1118 gagcagggggaatttaaccccttccctcccaaatccagtttgcgtgactatgttttaaa 1177
Db 15642 TTTTATTAATCATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 15583
Qy 1178 gaaacttttgcgttcaaatctcattacataacttttttttgcgcaaaaatttttgc 1237
Db 15582 TCATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 15523
Qy 1238 ttgctctcctgtgtaaaatttactgttttaggtactaactagccttgcgtgactgttc 1297
Db 15522 TTTTCTCTCTTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 15463
Qy 1298 tt 1299
Db 15462 TT 15461

RESULT 5
US-08-883-795A-36
Sequence 36, Application US/08883795A
Patent No. 5985607
GENERAL INFORMATION:
APPLICANT: Delcuve, Genevieve
APPLICANT: Awang, Gregor
TITLE OF INVENTION: Recombinant DNA Molecules and Expression
VECTORS FOR TISSUE PLASMINOGEN ACTIVATOR
NUMBER OF SEQUENCES: 39
CORRESPONDENCE ADDRESS:
ADDRESSEE: BERESKIN & PARR
STREET: 40 King Street West
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5H 3J2
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/883,795A
FILING DATE: 27-JUN-1997


```

; LENGTH: 3933 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genom
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: 198
US-08-731-722-3

```

Query Match 3.9%; Score 51.2; DB 4; Length 3933;
Best Local Similarity 54.3%; Pred. No. 0.076;
Matches 125; Conservative 0; Mismatches 103; Indels 2

53	Qy	actaatattttgggttaaataattaacacattattttttaagatcattaaattcaagaataataa	112
505	Db	ATTATTATTTCCTATTAAAAATTATGATGATTTTATTAAAGTAGAATAAGG--ATTACA	562
113	Qy	agattttttcaaaaaatgtataaaaattattatcatgatttttcatcacatttgatttt	172
563	Db	ATTTTCCTGTATAAAAAATAAAACCTTCTCTATGTGAAGAAATTTTTATAGCTACTGTTTA	622
173	Qy	gatacaaatatatttttttaattctctcaaaaaagtgcgaagacacttatagacata	232
623	Db	ATTAATAAAAGTTTATAAATTAACTCTTTTATAGACATTTAAAAAAAATAITATTAAAG	682
233	Qy	gtcttgctgtgttacaaagcattcatcatttaaacattaaaaatat	282
683	Db	GTAGAAATTCGAGAGAAATAAATACCTCTAAATTTACTTATTTATTAATAAATAT	732

RESULT

```

US-08-731-722-3/C
; Sequence 3, Application US/08731722
; Patent No. 5961971
; GENERAL INFORMATION:
; APPLICANT: Martin, Frank N.
; TITLE OF INVENTION: Biocontrol of Fungal S
; TITLE OF INVENTION: by Pythium oligandrum
; NUMBER OF SEQUENCE: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Saliwanchik & Saliwanchik
; SYREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: US
; ZIP: 32606-6669

```

Query Match 3.9%; Score 51.2; DB 4; Length 3933;
Best Local Similarity 54.3%; Pred. No. 0.076;
Matches 125; Conservative 0; Mismatches 103; Indels 2

Qy	53	actcaattatttgggttaaaataattaatcattatttttaagagatatataatgaagaattaaa	112
Db	3429	ATTATTATTTCCTTATTAATAATATTATGATTATTATTAAAGTAGAATAAGG--ATTACA	3372
Qy	113	agettctttcaaaaaatgtataaaaattattattcatgatcttttcacatcattgattttt	172
Db	3371	ATTTCTGTATAAANAATAAACCCTTCTATTGAGAAATTTTATAGCTACTGTTTA	3312
Qy	173	gatacaaatatatttttaatttcttaaaaaatgttgcagaacacctattagacata	232
Db	3311	ATTAAAAAGTTTATATATTTAAATCTTTATAGACATTTAAAAANAATATTATTAAATG	3252
Qy	233	gtcttgctctgtttcaaaaagcattcatcatttaatacatcaaaaaatat	282
Db	3251	GTAGAAATTCAGAGAATAAATACTTCTAATTTACTTTTATATTAATAATAT	3202

RESULT

US-08-883-795A-36/C
; Sequence 36, Application US/08883795A
; Patent No. 5985607
; GENERAL INFORMATION:
; APPLICANT: Delcuvie, Genevieve
; APPLICANT: Awang, Gregor
; TITLE OF INVENTION: Recombinant DNA Molecules and Expression
; TITLE OF INVENTION: Vectors for Tissue Plasminogen Activator
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BERESKIN & PARR
; STREET: 40 King Street West
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5H 3Y2
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS,
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/883.795A

Query Match 3.9%; Score 51; DB 4; Length 665;
Best Local Similarity 53.0%; Pred. No. 0.066;
Matches 131; Conservative 0; Mismatches 115; Indels

FILING DATE: 19-MAR-1992
ATTORNEY/AGENT INFORMATION:
NAME: No. 5773001is, Allen E.
REGISTRATION NUMBER: 34,490
REFERENCE/DOCKET NUMBER: 137-1061
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 354-3592
TELEFAX: (415) 857-1125
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 4570 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-764-100-20

[illegible]

RESULT 14
US-08-446-855A-1
Sequence 1, Application US/08446855A
Patent No. 5849573
GENERAL INFORMATION:
APPLICANT: Stewart, Thomas S
APPLICANT: Flores, Maria V
APPLICANT: O'Sullivan, William J
TITLE OF INVENTION: Nucleotide sequence
TITLE OF INVENTION: phosphate synthe
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Nixon & Vandertoye PC
STREET: 1100 No. 5849573th Giebe.R-
CITY: Arlington
STATE: Virginia
COUNTRY: USA
ZIP: 22201-4714
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM pc compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.24
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/446,855A
FILING DATE: 06-Jul-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Mitchard, Leonard C
REGISTRATION NUMBER: 29 009
REFERENCE/DOCKET NUMBER: 47-80
TELEPHONE INFORMATION:
TELEPHONE: 703-816-4000
TELEFAX: 703-816-4100
INFORMATION FOR SEQ ID NO: 1:

```

;
; SEQUENCE CHARACTERISTICS:
;
; LENGTH: 8920 base pair
;
; TYPE: nucleic acid
;
; STRANDEDNESS: single
;
; TOPOLOGY: linear
;
; MOLECULE TYPE: genomic
;
US-03-446-855A-1

```

	Query Match	3.8%	Score 50.2	DB 3	Length 8920
	Best Local Similarity	52.18	Pred. No. 0.13		
	Matches 135	Conservative 0	Mismatches 123	Indels 1	Gaps
QY	43	taccaatcatactaattattttgggttaaatatttaatactattttttaagatatttaatta	102		
Db	8491	TATAAATGATATATATAATAATAATAATAATAATAATAATAATAATAATAATAATA	8550		
QY	103	agaaattaaagaagattttttaaaaaaagtataaaaattatatttcgatgatttttccatc	162		
Db	8551	TTTAATTAATATTACGTTTAAATATATAATAATGTTTTATTAATAATGATCATTATTT	8610		
QY	163	atttgattttgataataaatatattttttttaattcttcaaaaagtgttcgaagacactt	222		
Db	8611	ATATTGATTATTATTTTATATAAATTTTGGTATATATACAAATTTTATTATTTCACAT	8670		
QY	223	at-tagacatagctctgtctgtttacaaaagcattcatcatttaatacatataaaaaata	281		
Db	8671	ATGTATAAACCCAAAATGGTTTTTTCAAATTTTCAANAATAATTTTAAATAAATTTA	8730		
QY	282	tttaatactaaacagtagaa	300		
Db	8731	TTAAATTATAAAAAAATAA	8749		

RESULT 15
US-08-565-907A-1/c
; Sequence 1, Application US/08565907A
; Patent No. 5814499
; GENERAL INFORMATION:
; APPLICANT: Sylvain Moineau, Barbara
; APPLICANT: J. Holler, Peter A. Vandenbergh,
; APPLICANT: Ebenezer R. Vedamuthu, Jeffrey K.
; APPLICANT: Kondo
; TITLE OF INVENTION: DNA Encoding Phage
; TITLE OF INVENTION: Abortive Infection Protein
; TITLE OF INVENTION: From Lactococcus
; TITLE OF INVENTION: lactis, and Method of Use Thereof
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ian C. McLeod
; STREET: 2190 Commons Parkway
; CITY: Okemos
; STATE: Michigan
; COUNTRY: USA
; ZIP: 48864
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette 5.25 inch,
; MEDIUM TYPE: 360 Kb storage
; COMPUTER: Acer
; OPERATING SYSTEM: MS-DOS (version 4)
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/565,907A
; FILING DATE: December 1, 1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Ian C. McLeod
; REGISTRATION NUMBER: 20,931
; REFERENCE/DOCKET NUMBER: Quest 4.1-152
; TELECOMMUNICATION INFORMATION:

TELEPHONE: (517) 347-4100
TELEFAX: (517) 347-4103
TELEX: NO. 581449e
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 4467
TYPE: Nucleotide
STRANDEDNESS: Single
TOPOLOGY: Linear
MOLECULE TYPE:
DESCRIPTION: Genomic DNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: N/A
ORIGINAL SOURCE:
ORGANISM: Lactococcus lactis
STRAIN:
INDIVIDUAL ISOLATE: W1
DEVELOPMENTAL STAGE: N/A
HAPLOTYPE: N/A
TISSUE TYPE: N/A
CELL TYPE: bacterium
CELL LINE: N/A
ORGANELLE: N/A
IMMEDIATE SOURCE:
LIBRARY: genomic
CLONE: SMQ-20
POSITION IN GENOME: N/A
FEATURE:
NAME/KEY: phage abortive infection
LOCATION: N/A
IDENTIFICATION METHOD: sequencing
OTHER INFORMATION: DNA encoding phage
OTHER INFORMATION: resistance
PUBLICATION INFORMATION: N/A
US-08-565-907A-1

Query Match 3.88; Score 49.8; DB 2; Length 4467;
Best Local Similarity 49.4%; Pred. No. 0.14; Mismatches 132; Indels 0; Gaps 0;
Matches 129; Conservative 0; Mismatches 132; Indels 0; Gaps 0;
Qy 51 atactaattatttgggttaataatcattatttttaagatatatttaagaataa 110
Db 2679 ATTTAAATCAATTAGTTAGTATTCTTCTGATTTTAAATCATCTATTTCAGAGAC 2620
Qy 111 aaagattttttaaaaaatgtataaaattatattattcattgatttttcacatttgatt 170
Db 2619 AACCATATACAAATATTTCATATAATTCCTTGATTAATAATTTTTCAGATAGTGTC 2560
Qy 171 ttgataataataatatt 230
Db 2559 TATTTTTCCTTTTAAAGCCCTTTGAATTTTACTAAATATTTTTCACATATTITGAAGC 2500
Qy 231 tagctctgtctgtttacaagcattcatcatttaataacatttaaaataatttaataact 290
Db 2499 TGATAAATCTGAAATCCAAATTCATTAATTTTCAAGAAAGTCAAAACTTATTAGT 2440
Qy 291 aacagtagaattctcttgga 311
Db 2439 TAATCTCAATCTTTTAATGA 2419

Search completed: September 2, 2000, 01:57:31
Job time: 8215 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 1, 2000, 23:33:51 ; Search time 161.21 seconds
(without alignments)
2039.280 Million cell updates/sec

Title: US-09-464-528-6
Perfect score: 1314
Sequence: 1 tctagatcaactcacatcc.....ttttgaagtataaccatgg 1314

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 311585 seqs, 125096042 residues

Total number of hits satisfying chosen parameters: 623170

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_36.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	65.8	5.0	6644	1 X33181	Base sequence of t
C 2	65.8	5.0	7372	1 X33182	Base sequence of t
C 3	65.8	5.0	7797	1 X33180	Cowpox virus bsr f
C 4	65.8	5.0	7996	1 X33184	Base sequence of t
C 5	60.2	4.6	9789	1 T41852	cDNA encoding Plas
C 6	56.4	4.3	5849	1 V33135	Plasmodium berghel
C 7	56	4.3	4590	1 N60472	Sequence encoding
C 8	56	4.3	19124	1 T72882	Plasmodium var-7 g
C 9	54.2	4.1	1186	1 V29580	Pythium oligandrum
C 10	53.6	4.1	19124	1 T72882	Plasmodium var-7 g
C 11	53.2	4.0	53585	1 X20251	Borrelia burgdorfe
C 12	52.6	4.0	605	1 T31530	Human 3' apolipop
C 13	52.4	4.0	26811	1 X20253	Borrelia burgdorfe
C 14	52.2	4.0	1826	1 V37413	Orpinomyces cellul
C 15	52.2	4.0	1826	1 V29477	Orpinomyces cellul
C 16	52.2	4.0	2503	1 Q53480	pNFX30 xylanase cD
C 17	52	4.0	8920	1 Q62924	Carbamoyl-phosphat
C 18	51.8	3.9	605	1 T31530	Human 3' apolipop
C 19	51.8	3.9	3975	1 N81157	Malaria-specific g
C 20	51.8	3.9	3975	1 Q22999	SERP gene. Recombi
C 21	51.8	3.9	6124	1 Q03568	Sequence encoding
C 22	51.8	3.9	26811	1 X20253	Borrelia burgdorfe
C 23	51.4	3.9	9789	1 T41852	cDNA encoding Plas
C 24	51.2	3.9	783	1 X20361	Borrelia burgdorfe
C 25	51.2	3.9	1711	1 V33136	Plasmodium berghel
C 26	51.2	3.9	1864	1 N71405	Sequence of ANS-1
C 27	51.2	3.9	3933	1 V29578	Pythium oligandrum
C 28	51.2	3.9	3933	1 V29578	Pythium oligandrum
C 29	51	3.9	1470	1 Q55185	MS-Lei610 vector.
C 30	50.8	3.9	602	1 T58577	Cryptosporidium pa
C 31	50.8	3.9	110000	1 V21209_11	Continuation (12 o
C 32	50.2	3.8	2418	1 Q27886	P.falciparum GBP13
C 33	50.2	3.8	4970	1 Q49959	Impatiens Necrotic

C 34 50.2 3.8 6124 1 Q03568 Sequence encoding
C 35 50.2 3.8 8920 1 Q62924 Carbamoyl-phosphat
C 36 50 3.8 2104 1 Q25273 Sequence encoding
C 37 49.8 3.8 4467 1 T68648 pNFX300 fragment 1
C 38 49.4 3.8 1611 1 Q12528 Thymidylate phosph
C 39 49.4 3.8 53585 1 X20251 Borrelia burgdorfe
C 40 49.2 3.7 1240 1 Q24177 Tox2a gene. DNA en
C 41 49.2 3.7 2503 1 Q53480 pNFX30 xylanase cD
C 42 49 3.7 1671 1 Q24134 50 kD subunit of S
C 43 49 3.7 4098 1 T43682 Medium chain-speci
C 44 48.8 3.7 6152 1 T78867 P. falciparum live
C 45 48.8 3.7 110000 1 V21209_12 Continuation (13 o

ALIGNMENTS

RESULT 1

X33181/c
ID X33181 standard; DNA; 6644 BP.
AC X33181,1999 (first entry)
DE Base sequence of the plasmid prx-lres-bar.
KW Cowpox virus; bsr; viral vector; expression; apoptosis; resistance;
KW crmA; bcl-2; bcl-x1; FLIP; survivin; IAP; adenovirus; cancer;
KW autoimmune disease; graft rejection reaction; inflammation;
KW inflammatory disease; ss.
OS Synthetic.
OS Cowpox virus.
PN W09913073-A2.
PD 18-MAR-1999.
PE 07-SEP-1998; J04010
PR 08-SEP-1997; JP-259235.
PA (RPRG-) RPR GENCELL ASIA PACIFIC INC.
PI Hamada H;
DI WPI; 99-243728/20.
PT New apoptosis-resistant virus-sensitive cell
PS Example 1; Page 38-41; 5lpp; English.
CC The present invention describes an apoptosis-resistant virus-sensitive cell line into which an apoptosis resistance gene has been introduced. The recombinant viruses generated are capable of expressing apoptosis-associated genes. These can then be used in a variety of diseases for which the induction of apoptosis by gene transfer, or where the inhibition of harmful apoptosis, is therapeutic. The recombinant viruses are useful as vectors for gene therapy which can be applied to cancer therapy for destroying cancer cells selectively, the treatment of autoimmune diseases and graft rejection reaction, and apoptosis induction therapy for inflammatory cells in inflammatory diseases. Prior arts have encountered the problem where if an adenovirus vector capable of expressing an apoptosis-associated gene is introduced into animal cells, the cells producing the virus will be destroyed because the period of time required to induce cell death by apoptosis is shorter than that required to replicate and produce the virus, resulting in failure to obtain a recombinant virus having the integrated apoptosis-associated gene. In this invention an apoptosis-resistant 293 cell line (having an apoptosis resistant gene introduced) is established and overcomes the problem. The present sequence represents the base sequence of the plasmid prx-lres-bsr, which contains the cowpox virus bsr gene, and is used in an example from the present invention.
SQ Sequence .6644 BP; 2166 A; 1573 C; 1424 G; 1481 T;

Query Match 5.0%; Score 65.8; DB 1; Length 6644;
Best Local Similarity 43.7%; Pred. NO. 0.0045;
Matches 289; Conservative 0; Mismatches 372; Indels 0; Gaps 0;

Oy 639 agttgagaccagacactggttcatactctctctctctctctctctctctctctctctc 698
Db 4442 AGCTTCGACCAATCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 4383
Oy 699 tcaagggtactttctctccctccctaccacactagatccgtgggtcaattcgatct 758
Db 4382 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 4323


```
FT /number= 2
FT /note= "no stop codon given"
PN WO9640766-A2.
PD 19-DEC-1996.
PF 07-JUN-1996; U09508.
PR 07-JUN-1996; US-487826.
PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.
PI Chitnis C, Miller LH, Peterson DS, Sim KL, Su X;
PI Wellens TE;
DR WPI: 97-052231/05.
DR P-PSDB; W22475.
PT New malaria vaccines - contains cysteine-rich DBL family protein
PT binding domains homologous domains of the Duffy and sialic acid
PT binding proteins
PS Claim 4; Page 56-61; 96pp; English.
CC This sequence represents the var-7 gene of Plasmodium. Var-7 belongs to
CC the Duffy binding like (DBL) family of genes which have homology to the
CC Duffy antigen binding protein (DABP) and sialic acid binding protein
CC (SABP) conserved regions (see T72889 and T72888 respectively). The var
CC family of genes modulate cytoadherence and antigenic variation of
CC Plasmodium infected erythrocytes. SABP and the Duffy antigen binding
CC protein (DABP) are soluble proteins that appear in the culture
CC supernatant after infected erythrocytes release merozoites. DABP and SABP
CC mediate the binding of merozoites and schizonts to the erythrocyte
CC surface. These proteins are necessary for erythrocyte invasion by the
CC parasite. This sequence can be used in the compositions of the invention.
CC The compositions are for the treatment and prevention of malaria, and
CC comprise either a nucleotide sequence or encoded polypeptide of the
CC var-1, var-2, var-3 or var-7 genes of the DBL gene family, a family of
CC genes having homology with conserved regions of DABP and SABP. The
CC compositions are used for the treatment and prevention of malaria. They
CC are also used in the preparation of vaccines for inducing a protective
CC immune response in a mammal to Plasmodium merozoites (especially
CC Plasmodium falciparum or Plasmodium vivax).
SQ Sequence 19124 BP; 7824 A; 2190 C; 2790 G; 6320 T;

Query Match 4.3%; Score 56; DB 1; Length 19124;
Best Local Similarity 47.7%; Pred. No. 0.17;
Matches 164; Conservative 0; Mismatches 180; Indels 0; Gaps 0;

QY 57 attatttgggttaataattcaattatttttaagatatttaataagaataaataagat 116
DB 1264 ATTATAATGTAATTAATTAATAAATATATTGTATAACATACAGACTAAAGAAAC 1323
QY 117 ttttttaaaaaatgtataaaattattattcattgatttttttcacattgatttgata 176
DB 1324 TATACAACTGCTACTAATAGTATATATATATATCTTTTATTATTAATGTCCT 1383
QY 177 atcaataatttttttttaattcttcaaaaaatgttgcaagacactatttagacatagt 236
DB 1384 CTTTTTTTTTTTTTTTAAATAAATAAATAATTAATATATATTTTTTTTCATAATATATGAT 1443
QY 237 tctctgttttcaaaagcattcatcatttaataacatttaaaaaataatttaactaacagt 296
DB 1444 TTAGTATTTTTATAATAAATAAATCTTTTAAAAAACTTCAAAACATTTTTCATATAAATA 1503
QY 297 aqaattcttcttgagtggtgagtaggcaactgagcattggaacagagagaagaaga 356
DB 1504 ATATTAAATTATTAGTACCACCTAGATAAATTAGAGAGAAACCTAGACATACCACAAAATA 1563
QY 357 gtcagacacagagacaaataaaaaagtgatgcacacaaataca 400
DB 1564 ATTAGAACAAGAAAGTATTACAAAAAATAAATAAATAAATAAATAA 1607

RESULT 9
V29580/c
ID V29580 standard; DNA; 1186 BP.
AC V29580;
DE 04-SEP-1998 (first entry)
KW Pythium oligandrum isolate 23-5 mitochondrial DNA partial sequence.
Pythium oligandrum; phytopathogenic; mitochondrial DNA; fungus;
```

```
KW soil saprophyte; pathogen; tomato; mycotoxic; plant protection;
KW vegetable crop; ss.
OS Pythium oligandrum.
FH Key Location/Qualifiers
FT misc_feature 502..684
FT /*tag= a
FT /note= "unique fragment claimed in claim 19"
PN WO9816110-A1.
PD 23-APR-1998.
PR 10-OCT-1997; U18343.
PR 17-OCT-1996; US-731722.
PA (UYFL ) UNIV FLORIDA.
PI Martin FN;
DR WPI: 98-250977/22.
PT Controlling phytopathogenic organisms with non-pathogenic Pythium
PT isolate - for control of damping off caused by Pythium
PS Claim 17; Page 30; 41pp; English.
CC This is a partial nucleotide sequence of the mitochondrial DNA from
CC a Pythium oligandrum isolate 23-5. Nucleotide sequences which are
CC inverted repeats, flanked by PstI restriction sites from mitochondrial
CC DNA from various P. oligandrum isolates are shown in V29576 to V29583.
CC These Pythium isolate sequences are non-pathogenic and can be used in a
CC method for controlling phytopathogenic organisms where the organisms are
CC contacted with such a Pythium isolate. The Pythium isolates are used to
CC control fungi, specifically pathogenic Pythium species (but possibly also
CC other soil-borne pathogens), particularly for protecting plants
CC (seedlings, transplants or vegetable crops such as tomato) against
CC damping off, especially after transplanting into open fields. The
CC isolate sequences are useful as sources of probes for identification of
CC particular isolates. The isolates are widely distributed in nature, with
CC a similar ecology to pathogenic species. They produce large quantities of
CC oospore inoculum on liquid or solid substrates and are tolerant of
CC several commonly used fungicides. A single application at the greenhouse
CC stage will protect plants after transplanting.
SQ Sequence 1186 BP; 469 A; 121 C; 125 G; 471 T;

Query Match 4.1%; Score 54.2; DB 1; Length 1186;
Best Local Similarity 59.3%; Pred. No. 0.33;
Matches 112; Conservative 0; Mismatches 73; Indels 4; Gaps 1;

QY 52 tactaatttttgggttaataattcaattatttttaagatatttaataagaataa 111
DB 595 TATTGAAGAATTTCTTATAACACTACTATAAGATTATAGGATCTACTGTATAAATCCAG 536
QY 112 aagatttttttaaaaaatgtataaaattattattcattgatttttcacattgatt 171
DB 535 ATCATATTATAAATAAATAAATAAAGTTT----TTCAATATTATATAAATAATATTC 480
QY 172 tgataataataatttttttttaatttttcaaaaaatgttgcaagacacttattgacat 231
DB 479 TTATTAGAAGTATTTCATTTTAAATTTTAAAAAGTTATATATATCTTTAAAAAGATAT 420
QY 232 agtcttggt 240
DB 419 AAATTAAT 411

RESULT 10
T72882/c
ID T72882 standard; cDNA; 19124 BP.
AC T72882;
DE 12-SEP-1997 (first entry)
DE Plasmodium var-7 gene.
KW DBL gene family; SABP; sialic acid binding protein; vaccine; therapy;
KW Duffy binding like gene; Duffy antigen binding protein; erythrocyte;
KW DABP; merozoite; malaria; var-1; var-2; var-3; var-7; immune response;
KW Plasmodium; ss.
OS Plasmodium falciparum.
FH Key Location/Qualifiers
FT exon 7317..15139
FT /*tag= a
```

FT Intron /number= 1
 FT 15140-16205
 FT /tag= b
 FT /number= 1
 FT 16206-17552
 FT /tag= c
 FT /number= 2
 FT /note= "no stop codon given"
 PN W09640766-A2.
 PD 19-DEC-1996.
 PR 07-JUN-1996; U09508.
 PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
 PI Chitnals C, Miller LH, Peterson DS, Sim KL, Su X;
 PI Wellens TE;
 DR WPI; 97-052231/05.
 DR P-PSDB; W22475.
 PT New malaria vaccines - contains cysteine-rich DBL family protein
 PT binding domains homologous domains of the Duffy and sialic acid
 PT binding proteins
 PS Claim 4; Page 56-61; 96pp; English.
 CC This sequence represents the var-7 gene of Plasmodium. Var-7 belongs to
 CC the Duffy binding like (DBL) family of genes which have homology to the
 CC Duffy antigen binding protein (DABP) and sialic acid binding protein
 CC (SABP) conserved regions (see T72889 and T72888 respectively). The var
 CC family of genes modulate erythrocyte adherence and antigenic variation of
 CC Plasmodium infected erythrocytes. SABP and the Duffy antigen binding
 CC protein (DABP) are soluble proteins that appear in the culture
 CC supernatant after infected erythrocytes release merozoites. DABP and SABP
 CC mediate the binding of merozoites and schizonts to the erythrocyte
 CC surface. These proteins are necessary for erythrocyte invasion by the
 CC parasite. This sequence can be used in the compositions of the invention.
 CC The compositions are for the treatment and prevention of malaria, and
 CC comprise either a nucleotide sequence or encoded polypeptide of the
 CC var-1, var-2, var-3 or var-7 genes of the DBL gene family, a family of
 CC genes having homology with conserved regions of DABP and SABP. The
 CC compositions are used for the treatment and prevention of malaria. They
 CC are also used in the preparation of vaccines for inducing a protective
 CC immune response in a mammal to Plasmodium merozoites (especially
 CC Plasmodium falciparum or Plasmodium vivax).
 CC Sequence 19124 BP; 7824 A; 2190 C; 2790 G; 6320 T;
 SQ

Query Match 4.1%; Score 53.6; DB 1; Length 19124;
 Best Local Similarity 45.9%; Pred. No. 0.41;
 Matches 221; Conservative 0; Mismatches 259; Indels 2; Gaps 1;

Qy 818 tgaacctactctcttcaataatcgcggaatacgcgttgagcttcagatcagatcg 877
 Db 15940 TTAATATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTA 15881
 Qy 878 aaatcatttcataatgccttcttcttcttcttcttcttcttcttcttcttctt 937
 Db 15880 AATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 15821
 Qy 938 tattcaaaaataaaccttgcgcgttgcgcgttgcgcgttgcgcgttgcgcgttgcgc 997
 Db 15820 TATTTT-AAATAAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 15763
 Qy 998 tttagcgaatttgcattgcattgcattgcattgcattgcattgcattgcattgcatt 1057
 Db 15762 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 15703
 Qy 1058 catactcccttagcttcaatttcttcttcttcttcttcttcttcttcttcttctt 1117
 Db 15702 TTTATGATATATATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 15643
 Qy 1118 gacgaggggaataatcccttcccttcccttcccttcccttcccttcccttccctt 1177
 Db 15642 TTTTATAATCATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 15583
 Qy 1178 gaaacttttgccttaaatcttattataaacttttatttatttatttatttatttatt 1237

Db 15582 TCATTTTATTTCTATCAAAATTTATATTTATTTATTTATTTATTTATTTATTT 15523
 Qy 1238 ttgctctctgttgtaaaatttactgtttaggtactactactagccttctgttgacgttt 1297
 Db 15522 TTTTCCTCTTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 15463
 Qy 1298 tt 1299
 Db 15462 TT 15461

RESULT 11
 X20251/c
 ID X20251 standard; DNA; 53585 BP.
 AC X20251;
 DE 04-MAY-1999 (first entry)
 DT Borrelia burgdorferi polynucleotide sequence #4.
 KW Borrelia burgdorferi; spirochete; bacterium; pathogen; Lyme disease;
 KW epidemic relapsing fever; endemic relapsing fever; Lyme borreliosis;
 KW infection; diagnosis; characterisation; detection; ds.
 OS Borrelia burgdorferi.
 PN W09858943-AL.
 PD 30-DEC-1998.
 PF 18-JUN-1998; U12764.
 PR 03-SEP-1997; US-057483.
 PR 20-JUN-1997; US-050359.
 PR 22-JUL-1997; US-053344.
 PR 22-JUL-1997; US-053377.
 PA (HUMA-) HUMAN GENOME SCI INC.
 PA (MEDI-) MEDIMUNE INC.
 PI Clayton R, Dougherty BA, Fraser C, Lathigra R, Smith HO,
 PI White OR;
 DR WPI; 99-081217/07.
 PT New Isolated Borrelia burgdorferi nucleic acids - used to develop
 PT products for the detection, diagnosis, characterisation, prevention
 PT and therapy of infections, particularly Lyme disease
 PS Claim 1; Page 801-831; 1128pp; English.
 CC X20248 to X20402 represent polynucleotide sequences isolated from
 CC Borrelia burgdorferi (fb). Products derived from Bb can be used for
 CC the detection, diagnosis, characterisation, prevention and therapy of
 CC Bb infections, e.g. Lyme disease. They can also be used for the
 CC production of biosynthetic products, e.g. enzymes. Borrelia belongs
 CC to a family of motile, spiral-shaped bacteria called Spirochetes.
 CC Spirochetes are pathogenic in humans and Borrelia causes epidemic and
 CC endemic relapsing fever, and Lyme borreliosis, more commonly known as
 CC Lyme disease.
 CC Sequence 53585 BP; 19235 A; 8081 C; 7001 G; 19266 T;
 SQ

Query Match 4.0%; Score 53.2; DB 1; Length 53585;
 Best Local Similarity 52.2%; Pred. No. 0.48;
 Matches 118; Conservative 0; Mismatches 108; Indels 0; Gaps 0;

Qy 21 aaacataacatggatctctcttaccatcatcatttatttgggttaaatattaac 80
 Db 33384 AAATATTAATAATATTAATAATATTTTAAACAATTAATTTTATTTATTAATAATG 33325
 Qy 81 attatttttaagatkatkaataaagaataaagaattttttaaataaataaataa 140
 Db 33324 CAAATTTTGTAAAAAATAAATAATTAATTTCAATCTTTTAAAGATTTTGAATAATTTT 33265
 Qy 141 tattattcatgatttttcatcatttgatttgataataataataataataataattct 200
 Db 33264 TTTTAAAGTTTATTTTGTGAAAAATATTTATTTGATGATGATTTATTTTATTA 33205
 Qy 201 taaaaaatgttgcagacacatttagacatgcttcttcttcttcttcttcttctt 246
 Db 33204 TTTTACCCTAAGGAGCTTATTTGAAAAACAGATTTTCTCTATAT 33159

RESULT 12
 T31530
 ID T31530 standard; cDNA; 605 BP.

Query Match 4.0%; Score 52.2; DB 1; Length 1826;
Best Local Similarity 49.8%; Pred. NO. 0.69;
Matches 132; Conservative 0; Mismatches 133; Indels 0; Gaps 0;

Search completed: September 2, 2000, 01:56:29
Job time: 8558 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 1, 2000, 23:32:16 ; Search time 5217.98 seconds
(without alignments)
449.425 Million cell updates/sec

Title: us-09-464-528-6
Perfect score: 1314
Sequence: 1 tctagataaacacacatcc.....ttttgaagtataaccatgg 1314

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 972840 seqs, 892348106 residues 1945680
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl:*

1: gb_ba1:*

2: gb_ba2:*

3: gb_ov:*

4: gb_ov:*

5: gb_pat:*

6: gb_ph:*

7: gb_pl1:*

8: gb_pl2:*

9: gb_pl3:*

10: gb_pr2:*

11: gb_pr3:*

12: gb_ro:*

13: gb_sts:*

14: gb_sy:*

15: gb_un:*

16: em_fun:*

17: em_hum1:*

18: em_hum2:*

19: em_in:*

20: em_or:*

21: em_or:*

22: em_ov:*

23: em_pat:*

24: em_ph:*

25: em_pl:*

26: em_ro:*

27: em_sts:*

28: em_sy:*

29: em_un:*

30: em_vl:*

31: gb_htg1:*

32: gb_htg2:*

33: gb_in1:*

34: gb_in2:*

35: em_ba1:*

36: em_ba2:*

37: em_hum3:*

38: em_hum4:*

39: gb_pr4:*

40: gb_htg3:*

41: gb_htg4:*

42: gb_htg5:*

43: gb_htg6:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	78	5.9	7218	5	I66494 Sequence 14
2	74	5.6	144759	10	HS352A20
C 3	70.2	5.3	153267	57	AC011212 Homo sapi
4	70	5.3	104992	41	AC005504 Plasmid
C 5	70	5.3	130281	60	AC004157 Plasmid
C 6	69.8	5.3	349919	54	AC008576 Homo sapi
C 7	69.6	5.3	164119	75	AC026640 Homo sapi
C 8	69.2	5.3	179310	71	AC013820 Homo sapi
9	69	5.3	48532	51	AC023371 Homo sapi
C 10	69	5.3	204951	60	AC005505 Plasmid
C 11	68.4	5.2	975	33	DDU87514 Dictyosteli
C 12	68.2	5.2	80920	60	AC006278 Plasmid
C 13	68.2	5.2	121024	8	CHMPX
C 14	67.8	5.2	910	13	CNS01G8P
15	67.4	5.1	256172	41	AC005139 Plasmid
C 16	67.2	5.1	43346	55	AC011556 Homo sapi
17	66.8	5.1	99263	32	CNS01DX9
18	66.8	5.1	152209	11	HS1108D11
19	66.6	5.1	1867	7	MTSCA323
20	66.4	5.1	149752	11	AC004616 Homo sapi
21	66.4	5.1	170427	41	AC006095 Homo sapi
C 22	66.4	5.1	183638	77	AC009653 Homo sapi
23	66.4	5.1	192581	31	PFMA113P1
24	66.4	5.1	216406	52	AC018919 Homo sapi

Submitted (07-MAY-1998) Sanger (ac.uk/hgp/Chr6/) Sanger Centre,
Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquerry@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On May 18, 1998 this sequence version replaced gi:2909620.

IMPORTANT: This sequence is the entire insert of clone 352A20.
During sequence assembly data is compared from overlapping clones
where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variations annotated may not be found in the sequence submission
corresponding to the overlapping clone as we submit sequences with
only a small overlap as described above.

This sequence was generated from part of bacterial clone contig
human chromosome 6, constructed by the Sanger Centre chromosome
mapping group. Further information can be found at
<http://www.sanger.ac.uk/HGP/Chr6/>

This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems
such as compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the
sequence is ambiguous, there is an annotation using the 'unsure'
feature key.

The true left end of clone 352A20 is at 1 in this sequence. The
true right end of clone 352A20 is at 144759.
352A20 is from the library RPc13 constructed at the Roswell Park
Cancer Institute by the group of Pieter de Jong.
For further details see <http://bacpac.med.buffalo.edu/>.

FEATURES

	Location/Qualifiers	source
repeat_region	1..144759 /organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="6" /map="q24.1-25.1" /clone="RP3-352A20" /clone_lib="RPCI-3"	
repeat_region	3..211 /note="AluSc repeat: matches 77..285 of consensus; incomplete repeat"	
repeat_region	334..392 /note="MER45 repeat: matches 4..62 of consensus"	
repeat_region	602..959 /note="AluSc repeat: matches 1..299 of consensus"	
repeat_region	1362..1544 /note="L1P83 repeat: matches 640..811 of consensus"	
repeat_region	1546..1836 /note="AluSg repeat: matches 299..1 of consensus"	
repeat_region	1839..1937 /note="L1MA4A repeat: matches 941..1046 of consensus"	
repeat_region	2448..2745	

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 22 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of 'N', but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1	1252:	contig of 1252 bp	in length
1253	1352:	gap of 100 bp	
1353	2580:	contig of 1198 bp	in length
2581	2650:	gap of 100 bp	
2651	4705:	contig of 2035 bp	in length
4706	4805:	gap of 100 bp	
4806	7966:	contig of 3161 bp	in length
7967	8066:	gap of 100 bp	
8067	10438:	contig of 2372 bp	in length
10439	10538:	gap of 100 bp	
10539	13144:	contig of 2586 bp	in length
13125	13224:	gap of 100 bp	
13225	15586:	contig of 2362 bp	in length
15587	15686:	gap of 100 bp	
15687	19054:	contig of 3368 bp	in length
19055	19154:	gap of 100 bp	
19155	23442:	contig of 4288 bp	in length
23443	23542:	gap of 100 bp	

*	23543	23542:	gap of	100 bp
*	23543	28034:	contig of	4492 bp in length
*	28035	28134:	gap of	100 bp
*	28135	31029:	contig of	2895 bp in length
*	31030	31129:	gap of	100 bp
*	31130	36055:	contig of	4896 bp in length
*	36026	36125:	gap of	100 bp
*	36126	42963:	contig of	6838 bp in length
*	42964	43063:	gap of	100 bp
*	43064	51794:	contig of	8731 bp in length
*	51795	51894:	gap of	100 bp
*	51895	60991:	contig of	9097 bp in length
*	60992	61091:	gap of	100 bp
*	61092	70105:	contig of	9014 bp in length
*	70106	70205:	gap of	100 bp
*	70206	77517:	contig of	7312 bp in length
*	77518	77617:	gap of	100 bp
*	77618	87726:	contig of	10109 bp in length
*	87727	87826:	gap of	100 bp
*	87827	100723:	contig of	12897 bp in length
*	100724	100823:	gap of	100 bp
*	100824	115535:	contig of	14712 bp in length
*	115536	115635:	gap of	100 bp
*	115636	134718:	contig of	19083 bp in length
*	134719	134818:	gap of	100 bp
*	134819	153267:	contig of	18449 bp in length

Location/Qualifiers
1. .153267

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="RP11-3N16"
/clone.lib="RPC1-11 Human Male BAC"
1..1952
/note="assembly fragment"

```

```

nature
1353. .2550
/notes="assembly_fragment"
2651. .4705
/notes="assembly_fragment"
4806. .7966
/notes="assembly_fragment"
8087. .10438
/notes="assembly_fragment"
10539. .13124
/notes="assembly_fragment"
13225. .15586
/notes="assembly_fragment"
15687. .19054
nature

```



```

TITLE      Plasmodium falciparum 3D7 chromosome 12
JOURNAL    Unpublished
REFERENCE  2 (bases 1 to 130281)
AUTHORS    Hyman, R.W., Qin, F., Fung, E.L., Conway, A.B. and Davis, R.W.
TITLE      Direct Submission
JOURNAL    Submitted (19-FEB-1998) Stanford DNA Sequencing and Technology
           Center, Stanford University, 855 California Avenue, Palo Alto, CA
           94304, USA
COMMENT    On Mar 15, 2000 this sequence version replaced gi:6652498.
           * NOTE: This is a 'working draft' sequence. It currently
           * consists of 3 contigs. The true order of the pieces
           * is not known and their order in this sequence record is
           * arbitrary. Gaps between the contigs are represented as
           * runs of N, but the exact sizes of the gaps are unknown.
           * This record will be updated with the finished sequence
           * as soon as it is available and the accession number will
           * be preserved.
           *
           * 1 67262: contig of 67262 bp in length
           * 67263 67462: gap of unknown length
           * 67463 82485: contig of 15023 bp in length
           * 82486 82685: gap of unknown length
           * 82686 130281: contig of 47596 bp in length.
           *
           * Location/Qualifiers
           *   1..130281
           *   /organism="Plasmodium falciparum"
           *   /db_xref="taxon:5833"
           *   /clone="3D7"
           *   /chromosome="12"
BASE COUNT 52250 a 11780 c 11855 g 53996 t 400 others
ORIGIN
Query Match      5.3%; Score 70; DB 60; Length 130281;
Best Local Similarity 57.0%; P-Id: No. 0.33;
Matches 147; Conservative 110; Indels 1; Gaps 1;
QY 53 actaatttttgggttaatttaattcatttttttaagatttaatttaagaataaa- 111
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 99223 ATTAATTATAATATATATATATATATATATATATATATATATATATATAT 99164
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 112 agatttttttaaaaaatgataaaattattattcatgatttttcacacatttgatt 171
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 99163 TTAATTATAATATATATATATATATATATATATATATATATATATATAT 99104
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 172 tgataataaatatttttttttaattcttaaaaaatgtgcgaagacatttagacat 231
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 99103 TAATTATAATATAAATTTAAATTTAAATTTAAATTTAAATTTAAATTTAAAT 99044
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 232 agcttggtctgtttacaagaagcattcatttaataacatttaaaaaattttaacta 291
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 99043 ATTTTATTATTATTAAATTTAAATTTAAATTTAAATTTAAATTTAAATTTAA 98984
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 292 acagtagaactctctgt 309
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 98983 AATAATAATATATATAT 98966

RESULT      6
AC008576/c  AC008576 349919 bp DNA HTG 18-FEB-2000
LOCUS       Homo sapiens chromosome 19 clone CTC-557J18, WORKING DRAFT
DEFINITION  SEQUENCE, 56 unordered pieces.
ACCESSION  AC008576
VERSION     AC008576.3 GI:6997021
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 349919)
AUTHORS     DOE Joint Genome Institute.
TITLE       Sequencing of Human Chromosome 19
JOURNAL     Unpublished

```

```

* 56312 59177: contig of 2866 bp in length
* gap of unknown length
* 59178 61728: contig of 2551 bp in length
* gap of unknown length
* 61729 65138: contig of 3410 bp in length
* gap of unknown length
* 65139 68509: contig of 3371 bp in length
* gap of unknown length
* 68510 72140: contig of 3631 bp in length
* gap of unknown length
* 72141 74177: contig of 2037 bp in length
* gap of unknown length
* 74178 76657: contig of 2480 bp in length
* gap of unknown length
* 76658 81120: contig of 4463 bp in length
* gap of unknown length
* 81121 83944: contig of 2824 bp in length
* gap of unknown length
* 83945 88969: contig of 5025 bp in length
* gap of unknown length
* 88970 94016: contig of 5047 bp in length
* gap of unknown length
* 94017 99820: contig of 5804 bp in length
* gap of unknown length
* 99821 103739: contig of 3919 bp in length
* gap of unknown length
* 103740 110294: contig of 6555 bp in length
* gap of unknown length
* 110295 116891: contig of 6597 bp in length
* gap of unknown length
* 116892 122399: contig of 5508 bp in length
* gap of unknown length
* 122400 129404: contig of 7005 bp in length
* gap of unknown length
* 129405 136213: contig of 6809 bp in length
* gap of unknown length
* 136214 142142: contig of 5929 bp in length
* gap of unknown length
* 142143 150401: contig of 8259 bp in length
* gap of unknown length
* 150402 158946: contig of 8545 bp in length
* gap of unknown length
* 158947 167955: contig of 9009 bp in length
* gap of unknown length
* 167956 177918: contig of 9963 bp in length
* gap of unknown length
* 177919 183990: contig of 6072 bp in length
* gap of unknown length
* 183991 191939: contig of 7949 bp in length
* gap of unknown length
* 191940 201055: contig of 9116 bp in length
* gap of unknown length
* 201056 210412: contig of 9357 bp in length
* gap of unknown length
* 210413 219973: contig of 9561 bp in length
* gap of unknown length
* 219974 228547: contig of 8574 bp in length
* gap of unknown length
* 228548 252971: contig of 24424 bp in length
* gap of unknown length
* 252972 284497: contig of 31526 bp in length
* gap of unknown length
* 284498 316714: contig of 32217 bp in length
* gap of unknown length
* 316715 349919: contig of 33205 bp in length.
* Location/Qualifiers
* 1..349919
*   /organism="Homo sapiens"
*   /db_xref="taxon:9606"
*   /chromosome="19"
*   /clone="CTC-557J18"
* 96739 a 81094 c 78642 g 92256 t 1188 others
* BASE COUNT
* ORIGIN

```

```

Query Match          5.3%; Score 69.8; DB 54; Length 349919;
Best Local Similarity 57.2%; Pred. No. 0.26; Mismatches 107; Indels 3; Gaps 1;
Matches 147; Conservative 0;

QY 53 actaatattttgggttaaatatcaatcaatttttaagatat---taattaagaatt 109
    || || || || || || || || || || || || || || || || || || ||
Db 309365 ACAACAATTCGGTTAAATGAAATGTTTATAAATATATATATTAATATAT 309306

QY 110 aagaatttttaaaaaatgataaattatattattattttttttttttttttgat 169
    || || || || || || || || || || || || || || || || || || ||
Db 309305 TATATATATATTAATATATATATATATATATATATATATATATATATAT 309246

QY 170 ttgataataaataatttttttttttttttttttttttttttttttttttttt 229
    || || || || || || || || || || || || || || || || || || ||
Db 309245 TATAATATATATATATATATATATATATATATATATATATATATATATAT 309186

QY 230 atagtccttctgtttacaaaagcattcatcttaataacataaaaaattttaac 289
    || || || || || || || || || || || || || || || || || || ||
Db 309185 TTATAATATATATATATATATATATATATATATATATATATATATATATAT 309126

QY 290 taacagtagaattctct 306
    || || || || || || || || || || || || || || || || || || ||
Db 309125 ATTTTATAATATATCTCT 309109

RESULT 7
AC026640
LOCUS AC026640 164119 bp DNA HTG 13-APR-2000
DEFINITION Homo sapiens chromosome 11 clone RP11-69K18 map 11, WORKING DRAFT
SEQUENCE, 16 unordered pieces.
ACCESSION AC026640
VERSION AC026640.2 GI:7547222
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 164119)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Brown,A., Burkett,G.,
Campopiano,A., Castle,A., Choepe,Y., Colangelo,M., Collins,S.,
Collamore,A., Cooke,P., DeArelano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Hearford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehotzky,J.,
Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
Meldrum,J., Meneus,L., Mihova,T., Miranda,C., Mienga,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neill,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (22-MAR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Apr 13, 2000 this sequence version replaced gi:7284664.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center

```


Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www.seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L8749

Center clone name: 69_K_18

----- Summary Statistics

Sequencing vector: M13; M77815; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 150998 bases at least Q40

Consensus quality: 158219 bases at least Q30

Consensus quality: 160955 bases at least Q20

Insert size: 176000; agarose-fp

Quality coverage: 3.8 in Q20 bases; agarose-fp

Quality coverage: 4.1 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently consists of 16 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

```

1 1608: contig of 1608 bp in length
* 1609 1708: gap of 100 bp
* 1709 3184: contig of 1476 bp in length
* 3185 3284: gap of 100 bp
* 3285 5541: contig of 2257 bp in length
* 5542 5641: gap of 100 bp
* 5642 7773: contig of 2132 bp in length
* 7774 7873: gap of 100 bp
* 7874 10690: contig of 2817 bp in length
* 10691 10790: gap of 100 bp
* 10791 13433: contig of 2643 bp in length
* 13434 13533: gap of 100 bp
* 13534 16455: contig of 2922 bp in length
* 16456 16555: gap of 100 bp
* 16556 21573: contig of 5018 bp in length
* 21574 21673: gap of 100 bp
* 21674 25705: contig of 4032 bp in length
* 25706 25805: gap of 100 bp
* 25806 31983: contig of 6178 bp in length
* 31984 32083: gap of 100 bp
* 32084 42802: contig of 10719 bp in length
* 42803 42902: gap of 100 bp
* 42903 54889: contig of 11987 bp in length
* 54890 54989: gap of 100 bp
* 54990 67682: contig of 12693 bp in length
* 67683 67782: gap of 100 bp
* 67783 80005: contig of 12223 bp in length
* 80006 80105: gap of 100 bp
* 80106 106699: contig of 26594 bp in length
* 106700 106799: gap of 100 bp
* 106800 164119: contig of 57320 bp in length.
Location/Qualifiers
1. 164119
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="11"
/map="11"
/clone="RP11-69K18"
1. 1608
/organism="Homo sapiens"
/clone="RP11-21P24"

```

FEATURES

```

misc_feature
misc_feature
misc_feature
misc_feature

```

```

/feature="assembly_fragment"
7874..10690
/feature="assembly_fragment"
10791..13433
/feature="assembly_fragment"
13534..16455
/feature="assembly_fragment"
16556..21573
/feature="assembly_fragment"
21674..25705
/feature="assembly_fragment"
25806..31983
/feature="assembly_fragment"
clone_end:SP6
vector_side:left
32084..42802
/feature="assembly_fragment"
42903..54889
/feature="assembly_fragment"
54990..67682
/feature="assembly_fragment"
67783..80005
/feature="assembly_fragment"
80106..106699
/feature="assembly_fragment"
106800..164119
/feature="assembly_fragment"

```

BASE COUNT 51462 a 29898 c 30038 g 51215 t 1506 others
ORIGIN

Query Match 5.3%; Score 69.6; DB 75; Length 164119;
Best Local Similarity 55.3%; Pred. No. 0.35;
Matches 135; Conservative 0; Mismatches 109; Indels 0; Gaps 0;

```

QY 68 ttaataataatcatatttttaagataataaataaagataaaatatttttaaaaaa 127
DB 118812 TAATATATATATATATATATATATATATATATATATATATATATATATTA 118871
QY 128 atgtataaataatattattattattattattattattattattattattatt 187
DB 118872 ATATATATATATATATATATATATATATATATATATATATATATATATA 118931
QY 188 tttttatttttaaaaaatgttgcaagacactatttagacatagcttctgtttta 247
DB 118932 TTAATTTTATATATATATATATATATATATATATATATATATATATATTA 118991
QY 248 caaagcattcatcatttaataacataaaaaataatttaataactaacagtagaattctt 307
DB 118992 ATATATATATATATATATATATATATATATATATATATATATATATATTA 119051
QY 308 gtga 311
DB 119052 GAGA 119055

```

RESULT 8
ACOL13820/c
LOCUS ACOL13820 179510 bp DNA HTG 01-APR-2000
DEFINITION Homo sapiens clone RP11-21P24, WORKING DRAFT SEQUENCE, 8 unordered
pieces.
ACCESSION ACOL13820
VERSION ACOL13820.3 GI:7382157
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 179510)
AUTHORS Birren, B., Linton, L., Nusbaum, C. and Lander, E.
TITLE Homo sapiens, clone RP11-21P24.
JOURNAL Unpublished.
REFERENCE 2 (bases 1 to 179510)

McEwan, P., McGurk, A., McKernan, K., McPheeters, R., Meldrim, J.,
 Meneus, L., Mihova, T., Miranda, C., Mlenga, V., Morrow, J., Naylor, J.,
 Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Olivari, T. M.,
 Peterson, K., Piere, N., Pisani, C., Pollara, V., Raymond, C.,
 Riley, R., Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S.,
 Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
 Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Tirrell, A.,
 Travers, M., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B.,
 Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zimmer, A. and
 Zody, M.
 Direct Submission
 Submitted (14-FEB-2000) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>
 ----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIGR
 Web site: <http://www-seq.wi.mit.edu>
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: L3985
 Center clone name: 21_D_18

NOTE: This record contains 55 individual
 * sequencing reads that have not been assembled into
 * contigs. Runs of N are used to separate the reads
 * and the order in which they appear is completely
 * arbitrary. Low-pass sequence sampling is useful for
 * identifying clones that may be gene-rich and allows
 * overlap relationships among clones to be deduced.
 * However, it should not be assumed that this clone
 * will be sequenced to completion. In the event that
 * the record is updated, the accession number will
 * be preserved.

1
 869: contig of 869 bp in length
 gap of unknown length
 870 1715: contig of 846 bp in length
 gap of unknown length
 1716 2597: contig of 882 bp in length
 gap of unknown length
 2598 3468: contig of 871 bp in length
 gap of unknown length
 3469 4369: contig of 901 bp in length
 gap of unknown length
 4370 5239: contig of 870 bp in length
 gap of unknown length
 5240 6103: contig of 864 bp in length
 gap of unknown length
 6104 6975: contig of 872 bp in length
 gap of unknown length
 6976 7871: contig of 896 bp in length
 gap of unknown length
 7872 8739: contig of 868 bp in length
 gap of unknown length
 8740 9613: contig of 874 bp in length
 gap of unknown length
 9614 10472: contig of 859 bp in length
 gap of unknown length
 10473 11328: contig of 856 bp in length
 gap of unknown length
 11329 12188: contig of 860 bp in length
 gap of unknown length
 12189 13069: contig of 881 bp in length
 gap of unknown length
 13070 13952: contig of 883 bp in length
 gap of unknown length
 13953 14829: contig of 877 bp in length
 gap of unknown length
 14830 15701: contig of 872 bp in length
 gap of unknown length
 15702 16587: contig of 886 bp in length
 gap of unknown length
 16588 17461: contig of 874 bp in length
 gap of unknown length
 17462 18326: contig of 865 bp in length
 gap of unknown length
 18327 19191: contig of 865 bp in length
 gap of unknown length
 19192 20081: contig of 890 bp in length
 gap of unknown length
 20082 20971: contig of 890 bp in length
 gap of unknown length
 20972 21849: contig of 878 bp in length
 gap of unknown length
 21850 22713: contig of 864 bp in length
 gap of unknown length
 22714 23562: contig of 849 bp in length
 gap of unknown length
 23563 24422: contig of 860 bp in length
 gap of unknown length
 24423 25286: contig of 864 bp in length
 gap of unknown length
 25287 26173: contig of 887 bp in length
 gap of unknown length
 26174 27055: contig of 882 bp in length
 gap of unknown length
 27056 28036: contig of 981 bp in length
 gap of unknown length
 28037 28916: contig of 880 bp in length
 gap of unknown length
 28917 29800: contig of 884 bp in length
 gap of unknown length
 29801 30670: contig of 870 bp in length
 gap of unknown length
 30671 31669: contig of 999 bp in length
 gap of unknown length
 31670 32564: contig of 895 bp in length
 gap of unknown length
 32565 33435: contig of 871 bp in length
 gap of unknown length
 33436 34363: contig of 928 bp in length
 gap of unknown length
 34364 35264: contig of 901 bp in length
 gap of unknown length
 35265 36131: contig of 867 bp in length
 gap of unknown length
 36132 37024: contig of 893 bp in length
 gap of unknown length
 37025 37922: contig of 898 bp in length
 gap of unknown length
 37923 38829: contig of 907 bp in length
 gap of unknown length
 38830 39709: contig of 880 bp in length
 gap of unknown length
 39710 40590: contig of 881 bp in length
 gap of unknown length
 40591 41490: contig of 900 bp in length
 gap of unknown length
 41491 42372: contig of 882 bp in length
 gap of unknown length
 42373 43253: contig of 881 bp in length
 gap of unknown length
 43254 44133: contig of 880 bp in length
 gap of unknown length
 44134 45009: contig of 876 bp in length
 gap of unknown length
 45010 45881: contig of 872 bp in length
 gap of unknown length
 45882 46757: contig of 876 bp in length
 gap of unknown length
 46758 47645: contig of 888 bp in length
 gap of unknown length
 47646 48532: contig of 887 bp in length.
 Location/Qualifiers

FEATURES

FEATURES	source	Location/Qualifiers
* 174761	177866:	contig. of 3106 bp in length
* 177867	178066:	gap of unknown length
* 178067	181356:	contig. of 3290 bp in length
* 181357	181556:	gap of unknown length
* 181557	184897:	contig. of 3341 bp in length
* 184898	185097:	gap of unknown length
* 185098	186708:	contig. of 1611 bp in length
* 186709	186908:	gap of unknown length
* 186909	190966:	contig. of 4058 bp in length
* 190967	191166:	gap of unknown length
* 191167	193608:	contig. of 2442 bp in length
* 193609	193808:	gap of unknown length
* 193809	198399:	contig. of 4591 bp in length
* 198400	198599:	gap of unknown length
* 198600	203091:	contig. of 4492 bp in length
* 203092	203291:	gap of unknown length
* 203292	204951:	contig. of 1660 bp in length.
BASE COUNT	83509 a	19825 c 24530 g 74582 t 2405 others
ORIGIN		
Query Match	5.3%	Score 69; DB '60; Length 204951;
Best Local Similarity	49.2%	Pred. No. 0.39; 215; Indels 1; Gaps 1;
Matches 209;	Conservative 0;	Mismatches 0;
Qy 884	tttcataatgctcttcttctttagcttatgagaaataaatacattttttttatttc 943	
Db 149838	TTTATTTAAATTAATTTTTTTTATTTTATTTAAATAAATTTTTTAAATTTTATTTA 149779	
Qy 944	aaataaaccttggcccttgctgactgagatggggcttggtgattcacagaattttgac 1003	
Db 149778	AATAAATTTTTTTTATTTTATTTTAAATAAATTTTTTTAGTTTTTTTATTTTAA 149720	
Qy 1004	gaatttgaattgactgttctgctgtagtttctgttcttctgttctgttcacataca 1063	
Db 149719	ATATTTTATTTTATGATATATATTTTATTTTAAATAATTTTTTTGCTTTTTTTT 149660	
Qy 1064	tctctaggcttcaattttatttcagtagtataggctcacataggaaattcgaagcag 1123	
Db 149659	TTTTATGATATATATTTTTTTTTTAAATAATATTTTTTCTTTTTTTTATTTATGATAT 149600	
Qy 1124	gggaattaaaccttcttccaaatccagtttggttgtatataatgtttaaaaaaagaaact 1183	
Db 149599	ATATTTTTTTTTTAAATGTTTTTTTTCTCTCTTTTTTATTTATTTTCAATAAATTT 149540	
Qy 1184	ttgctttaaattctataaacttttttatgcaaaaatttttgcagtgctttgct 1243	
Db 149539	TTTTTATTTTTTTTAAATCAATTTTTTTTATATATAAATAATTTTTTAAATTTCTTTTG 149480	
Qy 1244	ctcctgttgaaatttactgtttgaggtaactactaggcttggtagcagtttttgaag 1303	
Db 149479	ATAATATTTTTTTTATTTTATTTATCAAAATTTATATTTTTTAAATTTTATTTACT 149420	
Qy 1304	tataa 1308	
Db 149419	TTTAA 149415	
RESULT 11		
DDU87514/c		
LOCUS	975 bp	DNA
DEFINITION	Dictyostelium discoideum CAR3 gene, promoter region.	INV 05-APR-1997
ACCESSION	U87514	
VERSION	U87514.1	GI:1927211
KEYWORDS	Dictyostelium discoideum.	
ORGANISM	Dictyostelium discoideum	

REFERENCE AUTHORS TITLE	1 (bases 1 to 121024) Yamano,Y., Ohshima,K. and Komano,T. Nucleotide sequences of chloroplast 5S ribosomal RNA from cell suspension cultures of the liverworts Marchantia polymorpha and Jungermannia subulata Nucleic Acids Res. 12 (11), 4621-4624 (1984) 84247325	TITLE JOURNAL MEDLINE REFERENCE AUTHORS	Structure and organization of Marchantia polymorpha chloroplast genome. I. Cloning and gene identification J. Mol. Biol. 203 (2), 281-298 (1988) 89068685
JOURNAL MEDLINE REFERENCE AUTHORS TITLE	2 (bases 1 to 121024) Umesono,K., Inokuchi,H., Ohshima,K. and Ozeki,H. Nucleotide sequence of Marchantia polymorpha chloroplast DNA: a region possibly encoding three tRNAs and three proteins including a homologue of E. coli ribosomal protein S14 Nucleic Acids Res. 12 (24), 9551-9565 (1984) 85087956	TITLE JOURNAL MEDLINE REFERENCE AUTHORS	Structure and organization of Marchantia polymorpha chloroplast genome. III. Gene organization of the large single copy region from rbcL to trnI(CAU) J. Mol. Biol. 203 (2), 333-351 (1988) 89068687
JOURNAL MEDLINE REFERENCE AUTHORS TITLE	3 (bases 1 to 121024) Fukuzawa,H., Uchida,Y., Yamano,Y., Ohshima,K. and Komano,T. Molecular cloning of promoters functional in Escherichia coli from chloroplast DNA of a liverwort, Marchantia polymorpha Agric. Biol. Chem. 49, 2725-2731 (1985)	TITLE JOURNAL MEDLINE REFERENCE AUTHORS	Shimada,H. and Sugiyama,M. Fine structural features of the chloroplast genome: comparison of the sequenced chloroplast genomes Nucleic Acids Res. 19 (5), 983-995 (1991) 91212240
JOURNAL MEDLINE REFERENCE AUTHORS TITLE	4 (bases 1 to 121024) Yamano,Y., Kohchi,T., Fukuzawa,H., Ohshima,K. and Komano,T. Nucleotide sequences of chloroplast 4.5 S ribosomal RNA from a leafy liverwort, Jungermannia subulata, and a thalloid liverwort, Marchantia polymorpha FEBS Lett. 185, 203-207 (1985)	TITLE JOURNAL MEDLINE REFERENCE AUTHORS	Ohshima,K. Direct Submission Submitted (17-OCT-1986) K. Ohshima, Research Centre for Cell and Tissue culture, Faculty of Agriculture, Kyoto University, Kyoto 606, Japan GENERAL DESCRIPTION OF SEQUENCE gene name product name Ribosomal RNA genes 4.5S rRNA 5S rRNA 16S rRNA 23S rRNA
JOURNAL MEDLINE REFERENCE AUTHORS TITLE	5 (bases 1 to 121024) Ohshima,K., Fukuzawa,H., Kohchi,T., Shirai,H., Sano,T., Sano,S., Umesono,K., Shiki,Y., Takeuchi,M., Chang,Z., Aota,S., Inokuchi,H. Chloroplast gene organization deduced from complete sequence of liverwort Marchantia polymorpha chloroplast DNA Nature 322, 572-574 (1986)	COMMENT	rpobA: homologous to E.coli beta rpobB: homologous to E.coli beta' rpoc1: homologous to E.coli beta' rpoc2: homologous to E.coli Ribosomal protein genes and related genes 50S subunit rpl2: homologous to E. coli L2 L14 rpl14: homologous to E. coli L16 rpl16: homologous to E. coli L20 rpl20: homologous to E. coli L21 rpl21: homologous to E. coli L22 rpl22: homologous to E. coli L23 rpl23: homologous to E. coli L33 rpl33: homologous to E. coli 30S subunit rps2: homologous to E. coli S2 rps3: homologous to E. coli S3 rps4: homologous to E. coli S4 rps7: homologous to E. coli S7 rps8: homologous to E. coli S8 S11 rps11: homologous to E. coli S12 rps12: homologous to E. coli S14 rps14: homologous to E. coli S15 rps15: homologous to E. coli S18 rps18: homologous to E. coli S19 rps19: homologous to E. coli other genes homologous to E. coli Genes for photosynthesis rbcl: large subunit of RubisCO psaA: photosystem I P700 chlorophyll (an apoprotein) psaB: same as above psbA: photosystem II 32kd protein psbB: photosystem II P680 chlorophyll (an apoprotein) psbC: same as above psbD: photosystem II D2 protein psbE: cytochrome b559 psbF: same as above
JOURNAL MEDLINE REFERENCE AUTHORS TITLE	6 (bases 1 to 121024) Fukuzawa,H., Kohchi,T., Shirai,H., Ohshima,K., Umesono,K., Inokuchi,H. and Ozeki,H. Coding sequences for chloroplast ribosomal protein S12 from the liverwort, Marchantia polymorpha, are separated far apart on the different DNA strands FEBS Lett. 198, 11-15 (1986)		
JOURNAL MEDLINE REFERENCE AUTHORS TITLE	7 (bases 1 to 121024) Fukuzawa,H., Yoshida,T., Kohchi,T., Okumura,T., Sawano,Y. and Ohshima,K. Splicing of group II introns in mRNAs coding for cytochrome b6 and subunit IV in liverwort Marchantia polymorpha chloroplast genome: Exon specifying a region coding for two genes with the spacer region FEBS Lett. 220, 61-66 (1987)		
JOURNAL MEDLINE REFERENCE AUTHORS TITLE	8 (bases 1 to 121024) Kohchi,T., Ogura,Y., Umesono,K., Yamada,Y., Komano,T., Ozeki,H. and Ohshima,K. Ordered processing and splicing in a polycistronic transcript in liverwort chloroplasts Curr. Genet. 14 (2), 147-154 (1988) 89028845		
JOURNAL MEDLINE REFERENCE AUTHORS TITLE	9 (bases 1 to 121024) Kohchi,T., Shirai,H., Fukuzawa,H., Sano,T., Komano,T., Umesono,K., Inokuchi,H., Ozeki,H. and Ohshima,K. Structure and organization of Marchantia polymorpha chloroplast genome. IV. Inverted repeat and small single copy regions J. Mol. Biol. 203 (2), 353-372 (1988) 89068688		
JOURNAL MEDLINE REFERENCE AUTHORS TITLE	10 (bases 1 to 121024) Umesono,K., Inokuchi,H., Shiki,Y., Takeuchi,M., Chang,Z., Fukuzawa,H., Kohchi,T., Shirai,H., Ohshima,K. and Ozeki,H. Structure and organization of Marchantia polymorpha chloroplast genome. II. Gene organization of the large single copy region from rps'12 to atp8 J. Mol. Biol. 203 (2), 299-331 (1988) 89068686		
JOURNAL MEDLINE REFERENCE AUTHORS TITLE	11 (bases 1 to 121024) Ohshima,K., Fukuzawa,H., Kohchi,T., Sano,T., Sano,S., Shirai,H., Umesono,K., Shiki,Y., Takeuchi,M., Chang,Z., Aota,S., Inokuchi,H. and Ozeki,H.		

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 2, 2000, 00:24:41 ; Search time 2111.72 Seconds
(without alignments)
4521.091 Million cell updates/sec

Title: US-09-464-528-14
Perfect score: 2165
Sequence: 1 atcgatagacacatgttatt.....gttttgaagtataaagatg 2165

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 5247842 seqs, 2204914090 residues

Total number of hits satisfying chosen parameters: 10495684

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1: em_est1:*
2: em_est2:*
3: em_est3:*
4: em_est4:*
5: em_est5:*
6: em_est6:*
7: em_est7:*
8: em_est8:*
9: em_est9:*
10: em_est10:*
11: em_est11:*
12: em_est12:*
13: em_est13:*
14: em_est14:*
15: em_est15:*
16: em_est16:*
17: em_est17:*
18: em_est18:*
19: em_est19:*
20: gb_est1:*
21: gb_est2:*
22: gb_est3:*
23: gb_est4:*
24: gb_est5:*
25: gb_est6:*
26: gb_est7:*
27: gb_est8:*
28: gb_est9:*
29: gb_est10:*
30: gb_est11:*
31: gb_est12:*
32: gb_est13:*
33: gb_est14:*
34: gb_est15:*
35: gb_est16:*
36: gb_est17:*
37: gb_est18:*
38: gb_est19:*
39: gb_est20:*
40: gb_est21:*
41: gb_est22:*
42: gb_est23:*
43: gb_est24:*

44: gb_est25:*
45: gb_est26:*
46: gb_est27:*
47: gb_est28:*
48: gb_est29:*
49: gb_est30:*
50: gb_est31:*
51: gb_est32:*
52: em_est20:*
53: em_est21:*
54: em_est22:*
55: em_est23:*
56: em_est24:*
57: em_est25:*
58: em_est26:*
59: gb_est33:*
60: gb_est34:*
61: gb_est35:*
62: gb_est36:*
63: gb_est37:*
64: gb_est38:*
65: em_est27:*
66: em_est28:*
67: em_est29:*
68: em_est30:*
69: gb_est39:*
70: gb_est40:*
71: gb_est41:*
72: gb_est42:*
73: gb_est43:*
74: gb_est44:*
75: em_est31:*
76: em_est32:*
77: em_est33:*
78: em_est34:*
79: gb_est45:*
80: gb_est46:*
81: gb_est47:*
82: em_est35:*
83: em_est36:*
84: em_est37:*
85: gb_est48:*
86: gb_est49:*
87: gb_est50:*
88: gb_est51:*
89: gb_est52:*
90: gb_est53:*
91: gb_est54:*
92: gb_est55:*
93: gb_gss1:*
94: gb_gss2:*
95: gb_gss3:*
96: gb_gss4:*
97: em_gss1:*
98: em_gss2:*
99: em_gss3:*
100: em_gss4:*
101: gb_gss5:*
102: gb_gss6:*
103: gb_gss7:*
104: gb_gss8:*
105: gb_gss9:*
106: em_gss5:*
107: em_gss6:*
108: em_gss7:*
109: em_gss8:*
110: em_gss9:*
111: em_gss10:*
112: em_gss11:*
113: gb_gss10:*
114: gb_gss11:*
115: em_gss12:*
116: gb_gss12:*

pl and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES

Location/Qualifiers

1. .1101

/organism="Drosophila melanogaster"

/db_xref="taxon:7227"

/clone_lib="RPCI-98"

/clone="BACR08K10"

/note="end : TET3"

BASE COUNT 201 a 64 c 131 g 202 t 503 others

ORIGIN

Query Match 3.9%; Score 84; DB 122; Length 1101;
Best Local Similarity 18.5%; Pred. No. 4.6e-05;
Matches 129; Conservative 292; Mismatches 272; Indels 3; Gaps 1;

QY 249 aatcttatgtcttcacaaattagaacttgaaatttaatttaatttaaaactgaaacaa 308
DB 402 AMWWTTTTTTTTTAAWAAWAAATTAATTTAAWAAWAAATTTAAWAAWAAWATA 461
QY 309 atttgatcaattcatatcatagcttagtaataaaatgcgaataattgataaactc 368
DB 462 WTTTAAWAAWAAWAAWAAWAAWTTTWTWTTTAAWTTTAAWTTTAAWAAWAAA 521
QY 369 gcaaaagatttacaataatcttcgaaagaaataat---aacaataattgctgtttc 425
DB 522 AAAAAAATAAAWAAWAAWAAWTTTWTWTTTAAWAAWAAWAAWTTTWTWTTT 581
QY 426 atggttctgctgagagattgctgactatagaactctctcactgaccattcttgc 485
DB 582 YTYTWTWTTWTTWTTWTTWTTWTTWTTWTTWTTWTTWTTWTTWTTWTTWTTW 641
QY 486 ctcaactaaacagatgctgagattgctgagattgctgagattgctgagattgctgag 545
DB 642 YTYTWTWTTWTTWTTWTTWTTWTTWTTWTTWTTWTTWTTWTTWTTWTTWTTW 701
QY 546 aactctactactctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 605
DB 702 YTYTWTWTTWTTWTTWTTWTTWTTWTTWTTWTTWTTWTTWTTWTTWTTWTTW 761
QY 606 taacccttaacagtaaaattgaagagacaaataatatttttcaaaattgtagactat 665
DB 762 HAMWTHHAAWAAWAAWAAWTTTTHHHTTHHHTTHHHTTHHHTTHHHTTHHHT 821
QY 666 tttttttgtaattgacgaacaaacagatttctcctgatttctcctgatttctcctgac 725
DB 822 AYTCWTTWTTWTTWTTWTTWTTWTTWTTWTTWTTWTTWTTWTTWTTWTTWTT 881
QY 726 tgaactaaacaaattattatttttttcaaaacaaatttctcctgacgagctcctcag 785
DB 882 MHHWTHHHTTHHHTTHHHTTHHHTTHHHTTHHHTTHHHTTHHHTTHHHTTHH 941
QY 786 cccatgaaaaaaccttataaaattctacacattgacattgacattgacattgacatt 845
DB 942 MATWTTWTTWTTWTTWTTWTTWTTWTTWTTWTTWTTWTTWTTWTTWTTWTT 1001
QY 846 atgggtaaccagatcaaacctacacacacacacacacacacacacacacacacacac 905
DB 1002 WHYTWTWTTWTTWTTWTTWTTWTTWTTWTTWTTWTTWTTWTTWTTWTTWTT 1061
QY 906 ctaattatttttgggttaaatattattatttttttttttttttttttttttttttt 941
DB 1062 CTWYTHHTTHHHTTHHHTTHHHTTHHHTTHHHTTHHHTTHHHTTHHHTTHH 1097

RESULT 9

CNS00CS1/c

LOCUS

DEFINITION

BACR26H19 of RPCI-98 library from Drosophila melanogaster (fruit

CNS00CS1 843 bp DNA GSS 04-JUN-1999

Drosophila melanogaster genome survey sequence TET3 end of BAC #

BACR26H19 of RPCI-98 library from Drosophila melanogaster (fruit

fly), genomic survey sequence.

AL059666

AL059666.1 GI:4947129

GSS.

fruit fly.

Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;

Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 843)

Genoscope.

Direct Submission

Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :

Bp 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr

- Web : www.genoscope.cns.fr)

COMMENT

Determination of this BAC-end sequence was carried out as part of a

collaboration with the Berkeley Drosophila Genome Project (BDGP).

The BDGP is constructing a physical map of the Drosophila

melanogaster genome using these BACs. For further information

please see <http://www.fruitfly.org> The BDGP Drosophila

melanogaster BAC library was prepared by Kazutoyo Osoegawa and

Aaron Mammoser in Pieter de Jong's laboratory in the Department of

Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,

NY. The library is named RPCI-98 and was constructed by partial

EcoRI digestion of Drosophila DNA provided by the BDGP from the

isogenic strain v2: cn bw sp, the same strain used for the BDGP's

Pl and EST libraries. A more detailed description of the library

and how to order individual BAC clones, the entire library, or

filters for hybridization from the BACPAC Resource Center can be

found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES

Location/Qualifiers

1. .843

/organism="Drosophila melanogaster"

/db_xref="taxon:7227"

/clone_lib="RPCI-98"

/clone="BACR26H19"

/note="end : TET3"

BASE COUNT 131 a 102 c 259 g 141 t 210 others

ORIGIN

Query Match

Best Local Similarity 3.9%; Score 83.8; DB 122; Length 843;

Matches 122; Conservative 64; Mismatches 125; Indels 0; Gaps 0;

QY 851 taaccagatcaaacactccacataacataacatcgatattctctaccataactaat 910

DB 833 TATAWATAAAT 774

QY 911 tatttgggttaataatcaattatttttaagataatttaagataatttaagataatt 970

DB 773 TATTTATWTTTWTWTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 714

QY 971 tttaaaaaatgataaaattattattcattcattcattcattcattcattcattc 1030

DB 713 ATATTAT 654

QY 1031 aaatatatt 1090

DB 653 AWTWTWTWTTWTTWTTWTTWTTWTTWTTWTTWTTWTTWTTWTTWTTWTTWTT 594

QY 1091 ttctgtttcaaaagcattcattcattcattcattcattcattcattcattc 1150

DB 593 AAATWAT 534

QY 1151 aatcttcttctgt 1161

DB 533 WAATWATWAT 523

RESULT 10

CNS0021J/c

LOCUS

CNS0021J 1101 bp DNA GSS 03-JUN-1999

DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC # BACRO5N11 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.

ACCESSION AL061936

VERSION AL061936.1 GI:4940214

KEYWORDS GSS.

SOURCE fruit fly.

ORGANISM Drosophila melanogaster

REFERENCE Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 1101)

Genoscope.

Direct Submission

Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr. - Web : www.genoscope.cns.fr)

COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org/TheBDGP/Drosophila_melanogaster_BAC_library was prepared by Kazutoyo Osoegawa and Aaron Mammosier in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's pl and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES

source

1. .1101

/organism="Drosophila melanogaster"

/db_xref="taxon:7227"

/clone_lib="RPCI-98"

/clone="BACRO5N11"

/note="end : TET3"

BASE COUNT 631 a- 7 c 28 g 289 t 146 others

ORIGIN

Query Match 3.8%; Score 83; DB 122; Length 1101;

Best Local Similarity 40.2%; Pred. No. 6.7e-05;

Matches 210; Conservative 61; Mismatches 245; Indels 6; Gaps 1;

QY 740 tatttatttttttcaaaacaaatttcattgagcagcctccagccatgaaacac 799

DB 1043 TTTTCTTTTNNMAAATATTHMCACCTTTTCATMCTTCWTATTTTCMTMH 984

QY 800 ctataaaatattcacacattgaccattgaaagtcgtctcccatgggtacacagat 859

DB 983 TTTTAAATTTAAATTTATTTCTTACHATTYTTAACMCMWYTTACMCMWCAATC 924

QY 860 caaactcacatcccaatacatacattgattctccaccatacatacaattatt----- 914

DB 923 MCAWTTTAAWACTTAAACAAATATTAATTAATTAATTAATTAATTAATTAATTA 864

QY 915 -ttg99gttaaatattatttttttagattatttaataagaataataagaatttttt 973

DB 863 ATTTAAAAAAWTTTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWT 804

QY 974 aaaaaatgataaaattattattattattattattattattattattattattatt 1033

DB 803 ATTTAAAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 744

QY 1034 tataatt 1093

DB 743 TWTWTAAAAATTAATAATTTTAAAWAAAAAAATTAATAATTTTAAAAATTTTAT 684

QY 1094 tgtttcaaaagcattcatttattacattataaaataatttaataactaacagtagaat 1153

Db 683 WATWATAAAWATTTWTATTTTWTATTTTATTAATAAAWAAAAWATTTTAAAAW 624

QY 1154 ctcttctgagtggtggtgagtaggcaacctggcattgaaacagagagagagtcag 1213

Db 623 TTTTCTTTTAAWATTAATAATTAATAWATTTTATTAATTTTWTATTAATAA 564

QY 1214 accagaagacaaataaaatgtagcaacaaacaaacaaacaaacaaacaa 1255

Db 563 TTAWAAAAAAWAAAAAAWAAAAAAWAAAAAAWAAAAAAWAAAAWTA 522

RESULT 11

CNS016LI 1101 bp DNA GSS 26-JUL-1999

LOCUS Drosophila melanogaster genome survey sequence T7 end of BAC

DEFINITION BACN16D22 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.

ACCESSION AL106896

VERSION AL106896.1 GI:5624374

KEYWORDS GSS.

SOURCE fruit fly.

ORGANISM Drosophila melanogaster

REFERENCE Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 1101)

Genoscope.

Direct Submission

Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr. - Web : www.genoscope.cns.fr)

COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) - <http://www.edgp.ebi.ac.uk> - This Drosophila melanogaster BAC library (DrosBAC) was made by Alain Billaud at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBelobAC11.

FEATURES

Location/Qualifiers

source

1. .1101

/organism="Drosophila melanogaster"

/plasmid="pBelobAC11"

/db_xref="taxon:7227"

/clone_lib="DrosBAC"

/clone="BACN16D22"

/note="end : T7"

BASE COUNT 203 a 220 c 84 g 158 t 436 others

ORIGIN

Query Match 3.8%; Score 82.8; DB 123; Length 1101;

Best Local Similarity 33.6%; Pred. No. 7.2e-05;

Matches 113; Conservative 90; Mismatches 133; Indels 0; Gaps 0;

QY 923 aatattaatcattatttttaagatatattaagaataataagaattttttaaaaaatg 982

Db 1085 AWYWWYTTTAYWWYCTWTWTAATAAATAAATAAATAAATAAATAAATAAATA 1026

QY 983 tataaaattattattcattgatttttttttttttttttttttttttttttttttt 1042

Db 1025 TAATHTATWTYCAATAHAHWWWWMTTATATTAATTAATTAATTAATTAATTA 966

QY 1043 tttaattcttcaaaaatgttgcagacacattattagacataagcttcttcttttaca 1102

Db 965 WTYTYTATTAWAAHAAWATAATAATAATAATAATAATAATAATAATAATAATA 906

QY 1103 aagcattcatcatttataacattataataataataataataataataataata 1162

Db 905 WTAATATWATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 846

QY 1163 agtgggtggtgagtaggcaacctggcattgaaacagagagagagagagagagag 1222


```

QY 1127 aaaaatttt--aatactaacagtagaattcttctgtgagtggtggtgagtaggcaacc 1184
Db 788 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
QY 1185 tggcattgaacgagagagagagagtcagaaccagagacaaataaaaaagtatgcaacaa 1244
Db 848 TAWAWAWAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
QY 1245 acaaatcaaaa 1255
Db 908 MCWAATHAWAA 918

```

Search completed: September 2, 2000, 00:25:01
 Job time: 4325 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 2, 2000, 01:57:31 ; Search time 111.97 Seconds
(without alignments)
2659.716 Million cell updates/sec

Title: US-09-464-528-14

Perfect score: 2165

Sequence: 1 atcgatagacatgtatt.....gttttgaagtataaagatg 2165

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 243080 seqs, 68777915 residues

Total number of hits satisfying chosen parameters: 486160

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued_Patents_NA.*
1: /cgn2_6/ptodata/1/ina/5A.COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5B.COMB.seq.*
3: /cgn2_6/ptodata/1/ina/5C.COMB.seq.*
4: /cgn2_6/ptodata/1/ina/5D.COMB.seq.*
5: /cgn2_6/ptodata/1/ina/6.COMB.seq.*
6: /cgn2_6/ptodata/1/ina/PCTUS.COMB.seq.*
7: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	78	3.6	7218	1	US-08-232-463-14
2	57.2	2.6	3095	7	Sequence 14, Appl Patent No. 5231168
3	56	2.6	19124	4	US-08-487-826B-13
4	54.2	2.5	1186	4	Sequence 13, Appl
5	53.8	2.5	6243	4	Sequence 5, Appl
6	53.6	2.5	19124	4	US-09-056-075-1
7	53.2	2.5	3138	1	Sequence 1, Appl
8	52.6	2.4	665	4	US-08-487-826B-13
9	52	2.4	8920	3	Sequence 4, Appl
10	51.2	2.4	665	4	US-08-883-795A-36
11	51.2	2.4	3933	4	Sequence 1, Appl
12	51.2	2.4	3933	4	US-08-731-722-3
13	50.8	2.3	602	1	Sequence 3, Appl
14	50.6	2.3	8920	3	US-08-731-722-3
15	50.2	2.3	473	2	Sequence 1, Appl
16	50.2	2.3	4970	2	US-08-446-855A-1
17	50.2	2.3	4970	2	US-08-764-100-14
18	49.8	2.3	4467	2	Sequence 14, Appl
19	49.8	2.3	4467	2	US-08-764-100-20
20	49.8	2.3	4467	4	Sequence 20, Appl
21	49.4	2.3	1611	7	US-08-565-907A-1
22	49.2	2.3	319	1	Sequence 1, Appl
23	49	2.3	4098	4	US-08-910-551B-1
24	49	2.3	6768	2	Sequence 1, Appl
25	49	2.3	8457	1	US-08-909-425A-1
26	49	2.3	8457	4	Sequence 1, Appl
27	49	2.3	8457	4	US-07-991-867B-1
28	49	2.3	8457	4	US-08-544-332-1

27 48.4 2.2 1415 1 US-08-413-118-126 Sequence 126, App
28 48.4 2.2 1415 5 US-08-473-446-126 Sequence 126, App
29 48.2 2.2 6768 2 US-08-107-755A-1 Sequence 1, Appl
30 48.2 2.2 8457 1 US-07-991-867B-1 Sequence 1, Appl
31 48.2 2.2 8457 4 US-08-544-332-1 Sequence 1, Appl
32 47.8 2.2 12124 1 US-08-181-271A-36 Sequence 36, Appl
33 47.8 2.2 12124 1 US-08-449-315-36 Sequence 36, Appl
34 47.8 2.2 12124 1 US-08-444-803-36 Sequence 36, Appl
35 47.8 2.2 12124 1 US-08-449-043-36 Sequence 36, Appl
36 47.8 2.2 12124 2 US-08-456-265A-36 Sequence 36, Appl
37 47.8 2.2 12124 2 US-08-455-416-36 Sequence 36, Appl
38 47.8 2.2 12124 2 US-08-455-244-36 Sequence 36, Appl
39 47.8 2.2 12124 2 US-08-454-876-36 Sequence 36, Appl
40 47.8 2.2 12124 3 US-08-457-364-36 Sequence 36, Appl
41 47.8 2.2 12124 3 US-08-456-262-36 Sequence 36, Appl
42 47.8 2.2 12124 3 US-08-456-240-36 Sequence 36, Appl
43 47.8 2.2 12124 3 US-08-455-736-36 Sequence 36, Appl
44 47.8 2.2 12124 4 US-08-971-217-36 Sequence 36, Appl
45 47.6 2.2 5049 2 US-08-336-345-2 Sequence 2, Appl

ALIGNMENTS

RESULT 1
US-08-232-463-14
Sequence 14, Application US/08232463
Patent No. 5670367
GENERAL INFORMATION:
APPLICANT: DORNER, F.
APPLICANT: SCHEIFLINGER, F.
APPLICANT: FALKNER, F. G.
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 1800 Diagonal Road, Suite 500
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION NUMBER: US/08/232.463
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/935.313
FILING DATE:
APPLICATION NUMBER: EP 91 114 300.6
FILING DATE: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 30472/114 IMM
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)836-9300
TELEFAX: (703)683-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: ptz9pt-Fls
US-08-232-463-14

Qy	101	agacttaataaaattttacattttaattagaattttttttttatccaataataataatta	160
Db	2225	agaaatttaaaagatctcaaacatttgaatgtttatgttaatgcataatgagtg	2284
Qy	161	ttagttttattagaataatttaattagaataatttgaatcccgatttctctctcttct	220
Db	2285	ttattcaagtgtgaagtacaattaaaggaatagcaaggtatatagtataatt--tttcca	2342
Qy	221	tcgctatttcatttttbaaccacccaattttatgtttcttccaattagaaacttgaa	280
Db	2343	aaacattttaataattgaacaaaataaaaaaataaaaaaataataataataa	2402
Qy	281	attattaaataataaataactgaaacaaatttggatctcaattcatacacatgcttagta	340
Db	2403	atttttttctatttatgtgaactaatattatttattataataataataattattataa	2462
Qy	341	taaaatgcagataatttaattgataaaatctgcgaagagttttcaaatattctttcaaaaa	400
Db	2463	aagataactcaggaattctgtatataagaataaataatctgttatattattatagaataa	2522

Query Match 2.6%; Score 56; DB 4; Length 19124;
Best Local Similarity 47.7%; Pred. NO. 0.018;
Matches 164; Conservative 0; Mismatches 180; Indels 0; Gaps 0;


```

, REFERENCE/DOCKET NUMBER: 137-1061
,
, TELECOMMUNICATION INFORMATION:
, TELEPHONE: (415) 354-3592
,
, TELEFAX: (415) 857-1125
,
, INFORMATION FOR SEQ ID NO: 16:
, SEQUENCE CHARACTERISTICS:
, LENGTH: 473 base pairs
, TYPE: nucleic acid
, STRANDEDNESS: single
, TOPOLOGY: linear
,
, US-08-764-100-16

```

	Query Match	2.3%	Score 50.2;	DB 2;	Length 473;
	Best Local Similarity	52.1%	Pred. No. 0.12;		
	Matches 112;	Conservative	0;	Mismatches 103;	Indels 0; Gaps 0;
Qy	922	aaatattaacattatcttttaagatattaataaagaataataaagatttttttaaaaaaat	981		
Db	236	ANACCAAAAAAATTTTTTTTGTAATAATAAGCTCCGCCAGATTGGTCTAAGACC	295		
Qy	982	gtataaaattatatattctcatgatttttctacatatttgatttgataataaatattttt	1041		
Db	296	TTTTTATTGTGTTTTTATACATTTTTATTGTTTTTGTGATTATTATTATTATTATTA	355		
Qy	1042	ttttaattcttcaaaaaatgttcgaacacctattatagacatagctctgtctctgttaca	1101		
Db	356	TATTTTTTTATATAGTTTGGCTTATTTTAACACTTATTATAGACAAATAAATTATTGTGATTA	415		
Qy	1102	aagcattcatcatcttataatcaattaaaaaatattt	1136		
Db	416	CAATCATCTCGCTTATTATTATTATTTAAACACATTT	450		

Search completed: September 2, 2000, 01:59:26
Job time: 8330 sec

Cowpox virus; bsr: viral vector; expression; apoptosis; resistance; crma; bcl-2; bcl-xl; FLIP; survivin; IAP; ILP; adenovirus; cancer; autoimmune disease; graft rejection reaction; inflammation; inflammatory disease; ss.

Cowpox virus.
W09913073-A2.
PN
PD 18-MAR-1999.
PF 07-SEP-1998; J04010.
PR 08-SEP-1997; JP-259235.
PA (RPRG-) RPR GENCELL ASIA PACIFIC INC.
PI Hamada H;
PT WPI: 99-243728/20.
New apoptosis-resistant virus-sensitive cell

Example 1; Page 34-38; Sllp: English.
CC The present invention describes an apoptosis-resistant virus-sensitive cell line into which an apoptosis resistance gene has been introduced. The recombinant viruses generated are capable of expressing apoptosis-associated genes. These can then be used in a variety of diseases for which the induction of apoptosis by gene transfer, or where the inhibition of harmful apoptosis, is therapeutic. The recombinant viruses are useful as vectors for gene therapy which can be applied to cancer therapy for destroying cancer cells selectively, the treatment of autoimmune diseases and graft rejection reaction, and apoptosis induction therapy for inflammatory calls in inflammatory diseases. Prior arts have encountered the problem where if an adenovirus vector capable of expressing an apoptosis-associated gene is introduced into animal cells, the cells producing the virus will be destroyed because the period of time required to induce cell death by apoptosis is shorter than that required to replicate and produce the virus, resulting in failure to obtain a recombinant virus having the integrated apoptosis-associated gene. In this invention an apoptosis-resistant 293 cell line (having an apoptosis resistant gene introduced) is established and overcomes the problem. The present sequence represents the cowpox virus bsr gene which is used in an example from the present invention.

Sequence 7797 BP; 2542 A; 1760 C; 1656 G; 1839 T;

Query Match 3.0%; Score 65.8; DB 1; Length 7797;
Best Local Similarity 43.7%; Pred. No. 0.0084;
Matches 289; Conservative 0; Mismatches 372; Indels 0; Gaps 0

QY 1491 agttagagacaagaacacactgcttcatactcctctgtctctctctctctctctcacc 1550
DB |||||
DB 5595 AGCTCGACCAATCTTTT|||||TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 5336
QY 1551 tcaaggtaacctttctctccctaccacaaactagattccgtgggtccaattcggaatc 1610
DB |||||
DB 5535 TTTT|||||TT 5476
QY 1611 tgcaactctggcttgcttgctctcttcccaactgggtccactcaggatccatgt 1670
DB |||||
DB 5475 TTTT|||||TT 5416
QY 1671 gaaactctactcttctttaatactgcggaatacgcgttgagacttcagactcagtcga 1730
DB |||||
DB 5415 TTTT|||||TT 5356
QY 1731 atcatctcataatgccccttctctctttagcttatgagaaaataaatctttttttt 1790
DB |||||
DB 5355 TTTT|||||TT 5296
QY 1791 atttcaaaaataaaccttgggcccttgtcgtgactgagatgggggttggtgattcacagaatt 1850
DB |||||
DB 5295 TTTT|||||TT 5236
QY 1851 ttacgcaattttgtaattgacttgttctgtctgtagtgtttgtttctctctgtctctc 1910
DB |||||
DB 5235 TTTT|||||TT 5176
QY 1911 atacattcttagcttccaattttattccagatagtaggtccacaaatcaggaattcaaacttgg 1970
DB |||||
DB 5175 TTTT|||||TT 5116

Qy	1971	agcagggggaataatcccttcctcaaaaccacggtttgttgtatatatgtttaaanaagt	2030
Db	5115	TTT	5056
Qy	2031	aacttttgcttaaatcatcataaacatttttttataggcaaaaatttttgcagtgtct	2090
Db	5055	TTT	4996
Qy	2091	tgtctctcgttgtaaaattactgttaggtactaacctcaggctgttgtgcagtttt	2150
Db	4995	TTT	4936
Qy	2151	t 2151	
Db	4935	T 4935	
RESULT 4			
X	33184/c		
ID	X33184 standard; DNA; 7996 BP.		
AC	X33184;		
DT	25-JUN-1999 (first entry)		
DE	Base sequence of the plasmid prx-Bcl 2-i-hcd 25.		
KW	Coxov virus; bsr; viral vector; expression; apoptosis; resistance;		
KW	cma; bcl-2; bcl-xl; FLIP; survivin; IAP; ILP; adenovirus; cancer;		
KW	autoimmune disease; graft rejection reaction; inflammation;		
KW	inflammatory disease; ss.		
OS	Synthetic.		
OS	Homo sapiens.		
PN	W0913073-A2.		
PF	18-MAR-1999.		
PD	07-SEP-1998; J04010.		
PR	08-SEP-1997; JP-259235.		
PA	(RPRG-) RPR GENCELL ASIA PACIFIC INC.		
PI	Hamada H;		
DR	WPI: 99-243728/20.		
PT	New apoptosis-resistant virus-sensitive cell		
PS	Example 3; Page 46-49; Sipp; English.		
CC	The present invention describes an apoptosis-resistant virus-sensitive		
CC	cell line into which an apoptosis resistance gene has been introduced.		
CC	The recombinant viruses generated are capable of expressing apoptosis-		
CC	associated genes. These can then be used in a variety of diseases for		
CC	which the induction of apoptosis by gene transfer, or where the		
CC	inhibition of harmful apoptosis, is therapeutic. The recombinant virus		
CC	are useful as vectors for gene therapy which can be applied to cancer		
CC	therapy for destroying cancer cells selectively, the treatment of		
CC	autoimmune diseases and graft rejection reaction, and apoptosis induc-		
CC	therapy for inflammatory cells in inflammatory diseases. Prior arts ha-		
CC	ve encountered the problem where if an adenovirus vector capable of		
CC	expressing an apoptosis-associated gene is introduced into animal cell		
CC	the cells producing the virus will be destroyed because the period of		
CC	time required to induce cell death by apoptosis is shorter than that		
CC	required to replicate and produce the virus, resulting in failure to		
CC	obtain a recombinant virus having the integrated apoptosis-associated		
CC	gene. In this invention an apoptosis-resistant 293 cell line (having a		
CC	apoptosis resistant gene introduced) is established and overcomes the		
CC	problem. The present sequence represents the base sequence of the		
CC	plasmid prx-Bcl 2-i-hcd 25, which contains the human Bcl-2 gene, and		
CC	is used in an example from the present invention.		
SO	Sequence 7996 BP: 2463 A: 1829 G: 1689 T:		

```

Query Match      3.0%;  score 65.8;  DB 1;  Length 7996;
Best Local Similarity 43.7%;  Pred. No. 0.0084;
Matches 289;  Conservative 0;  Mismatches 372;  Indels 0;  Gaps 0;

Qy 1491 agtggaccagaacacactggttcatatctctctctctctctctctctctctctc 1550
      ||| ||||| | | | | | | | | | | | | | | | | | | | |
Db 5794 AGCTCGACCAACTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 5735

Qy 1551 tcagggtactttctctccctctaccacaaactcagatctcgggtcaatttcgggatct 1610
      ||| ||||| | | | | | | | | | | | | | | | | | | | |
Db 5734 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 5675

```


CC the detection, diagnosis, characterisation, prevention and therapy of
CC Bb infections, e.g. Lyme disease. They can also be used for the
CC production of biosynthetic products, e.g. enzymes. Borrelia belongs
CC to a family of motile, spiral-shaped bacteria called Spirochetes.
CC Spirochetes are pathogenic in humans and Borrelia causes epidemic and
CC endemic relapsing fever, and Lyme borreliosis, more commonly known as
CC Lyme disease.
SQ Sequence 116277 BP; 42656 A; 19868 C; 14490 G; 39250 T;

Query Match 2.7%; Score 58; DB 1; Length 116277;
Best Local Similarity 48.8%; Pred. No. 0.15;
Matches 157; Conservative 0; Mismatches 165; Indels 0; Gaps 0;
QY 65 tatctgtttatgatcttcaggcgcaaaaatgcgagctacttaataaaattttacattt 124
DB 18742 TCTCTTCTGTTTCAATTTGCTTATAAATTTTGTAACTTTCAGAAAAATTTAAATTT 18683
QY 125 aaattagaattttttatcaataaatttaattatttagttttttattgagaataattatt 184
DB 18682 TGAATTTTCTTCTTCATTTAAATATATCCATGGATTTTGTTCATATTAATAAGA 18623
QY 185 agaaaaatttgatcccgatttcctcccttcttctgctattcatcttcttaacca 244
DB 18622 CTTCTGTTTCAATTTTAACTTTTAAAGTGTGTACAAAAATAAATTTATTTA 18563
QY 245 aaccaattcttatgtcttccaaattagaacttgaaattatttaattataataaaactgaa 304
DB 18562 TTGTAACCTTACTTTTAAATTAATGATTAATAATTAAGGGAGAAATTTTATGTA 18503
QY 305 acaatttggtatcaattcaatacatcgttagtaataaaatgcgataatttaattgataa 364
DB 18502 TAAATATGTTTTTTTAAAACTATTGTTCATTTGTTTAAATTTTGTAGTAATGCTTG 18443
QY 365 atctgcaaaagattttacaatt 386
DB 18442 TACTTCAAAAGATAGCTCAAT 18421

RESULT 12
Q03875
ID Q03875 standard; DNA: 3095 BP.
AC Q03875;
DE 24-AUG-1990 (first entry)
KW Sequence encoding carboxylic terminal part of native GLURP.
OS Plasmodium falciparum; antigen; malaria; vaccine; GLURP; ss.
FH Plasmodium falciparum.
FT Key Location/Qualifiers
FT cds 1..2352
FT /tag= a
FT /product=GLURP
FT PN W09022811-A.
PD 22-MAR-1990.
PF 18-SEP-1989; 00218.
PR 03-MAR-1989; US-218885.
PR 03-MAR-1989; DK-005191.
PA (STAT-) Statens Seruminst.
PI Dzilegiel M, Borre M, Jepsen S, Vuust J, Rieneck K, Wind A, Jakobsen PH;
DR WPI: 90-115998/15.
DR P-PSDB: R05804.
PT Polypeptide(s) derived from Plasmodium falciparum antigen - used in
PT vaccines and in production of antibodies, for diagnosis and
PT therapy of malaria.
PS Disclosure; Fig 7: 108pp; English.
CC An open reading frame of 2349 bps extends from the 5' terminal end of the
CC sequence displays some of the characteristics of other malaria nucleic
CC acid sequences: tandemly repeated motifs, high AT content and a
CC corresponding preference for codons containing these bases, and a high
CC content of codons for glutamate. Three major repetitive sequences are:
CC one motif from bp 34 to bp 156 is repeated from bp 289 to bp 411; another
CC motif from bp 477 to bp 521 is repeated tandemly twice from bp 522 to bp
CC 566 and from bp 567 to bp 611; a third motif from bp 1174 to bp 1233 is

CC repeated tandemly 11 times. This last repetitive region consists of 360bp
CC repeats differing only in 3 bases GAR coding for aspartate. This region
CC is flanked to the 5' terminal of a degenerated 60 bp repeat. GC content
CC of the coding part of the insert is on average 30%, and of the non-coding
CC 3' terminal 11%.
SQ Sequence 3095 BP; 1443 A; 300 C; 491 G; 861 T;

Query Match 2.6%; Score 57.2; DB 1; Length 3095;
Best Local Similarity 50.3%; Pred. No. 0.19;
Matches 167; Conservative 0; Mismatches 163; Indels 2; Gaps 1;
QY 101 agtacttaataaaattttacatttaaaattagaattttttttatcaataaaatttaattta 160
DB 2225 AGAAATTTAAAAAGATATCTCAAACTATTGTAAAGTGTATGATTAATCATATGATGTG 2284
QY 161 ttagtctttatagaaatttaattagaataatttgaatccccgatttctctccctttctt 220
DB 2285 TTATTCAGTTGTAGTACANTTAAGGAANTAGCAAGATATAGTAATATT--TTTCCA 2342
QY 221 tegtattcatcattttctaaccaaaccaacttttatgtttctcaaataggaaacttgaa 280
DB 2343 AAACATTTAAATAATTAACAAAAAATAAATAAATAAATAAATAAATAAATAAATAA 2402
QY 281 attattattataataaactgaacaaactttgggtatcaattcatatcatatcattagtaa 340
DB 2403 ATTTTCTTCTTATATATGTAACATAATTTATTTATTAATAAATATATATATATATAA 2462
QY 341 laaaatgcgataaataaattgatacaactctgcaaaagattttcaaatatcttccagaaaa 400
DB 2463 AAGAATACTAGGATTTGGTGTATATATAGAAATAAATAATTCATTGTATATATAGAAAA 2522
QY 401 attataaacaattttgtcgttttctcatggtgt 432
DB 2523 AATAAAATAAATAATTTCTTTCTTAGTTGT 2554

RESULT 13
Q03568
ID Q03568 standard; DNA: 6124 BP.
AC Q03568;
DE 02-AUG-1990 (first entry)
KW Sequence encoding the SERA protein of Plasmodium.
OS Malaria; SERA protein; vaccine; ds.
FH Plasmodium falciparum.
FT Key Location/Qualifiers
FT misc_feature 1..6
FT /tag= a
FT /note= "Sequence obscured."
FT misc_feature 121..128
FT /tag= b
FT /note= "Sequence obscured."
FT misc_feature 241..245
FT /tag= c
FT /note= "Sequence obscured."
FT exon 2407..2440
FT /tag= d
FT intron 2441..2598
FT /tag= e
FT exon 2599..3405
FT /tag= f
FT intron 3406..3580
FT /tag= g
FT exon 3581..3721
FT /tag= h
FT intron 3722..3850
FT /tag= i
FT exon 3851..5835
FT /tag= j
FT allele 3087
FT /tag= k
FT allele 3092
FT /label=Allele.II.

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 2, 2000, 02:13:13 ; Search time 5217.98 Seconds
(without alignments)
740.491 Million cell updates/sec

Title: US-09-464-528-14
Perfect score: 2165
Sequence: 1 atcgatagacatgtatt.....gttttgaagtataaagatg 2165

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 972840 seqs, 892348106 residues

Total number of hits satisfying chosen parameters: 1945680

Minimum DB seq length: 0
Maximum DB seq length: 1000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl:
1: gb_ba1:
2: gb_ba2:
3: gb_om:
4: gb_ov:
5: gb_pat:
6: gb_ph:
7: gb_pl1:
8: gb_pl2:
9: gb_pr1:
10: gb_pr2:
11: gb_pr3:
12: gb_ro:
13: gb_sts:
14: gb_sy:
15: gb_un:
16: em_fun:
17: em_hum1:
18: em_hum2:
19: em_in:
20: em_om:
21: em_or:
22: em_ov:
23: em_pat:
24: em_ph:
25: em_pl:
26: em_ro:
27: em_sts:
28: em_sy:
29: em_un:
30: em_vl:
31: gb_htg1:
32: gb_htg2:
33: gb_in1:
34: gb_in2:
35: em_ba1:
36: em_ba2:
37: em_hum3:
38: em_hum4:
39: gb_pr4:
40: gb_htg3:
41: gb_htg4:
42: gb_htg5:
43: gb_htg6:

44: gb_htg7:
45: em_htg1:
46: em_htg2:
47: em_htg3:
48: em_hum5:
49: gb_pl3:
50: gb_pr5:
51: gb_htg8:
52: gb_htg9:
53: gb_htg10:
54: gb_htg11:
55: gb_htg12:
56: gb_htg13:
57: gb_htg14:
58: gb_in3:
59: gb_htg15:
60: gb_htg16:
61: gb_htg17:
62: em_htg4:
63: em_htg5:
64: em_htg6:
65: em_htg7:
66: em_hum6:
67: gb_htg18:
68: gb_htg19:
69: gb_htg20:
70: gb_htg21:
71: gb_htg22:
72: gb_htg23:
73: gb_htg24:
74: gb_htg25:
75: gb_htg26:
76: gb_htg27:
77: gb_htg28:
78: gb_htg29:
79: gb_htg30:
80: gb_htg31:
81: gb_vil:
82: gb_vl2:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	118.4	5.5	4601	33	DMU11584
2	118.4	5.5	19517	58	DMU37541
3	99.8	4.6	86829	33	PFMAL3P5
4	98.4	4.5	86829	33	PFMAL3P5
5	97.4	4.5	14867	34	AE001398
6	92.6	4.3	146285	39	AC005083
7	91.8	4.2	67970	33	PFMAL1P3
8	91.2	4.2	153098	33	PFMAL3P2
9	91.2	4.2	5371	7	YSCMTTGA
10	89.6	4.1	67970	33	PFMAL1P3
11	89.6	4.1	104992	41	AC003504
12	89.6	4.1	130281	60	AC004157
13	89	4.1	161891	55	AC008206
14	88.8	4.1	253307	33	PFMAL3P7
15	88.2	4.1	80518	31	PFMAL13PA
16	88	4.1	170125	41	AC007465
17	87.4	4.0	4601	33	DMU37541
18	87.4	4.0	19517	58	DMU37541
19	86.8	4.0	104992	41	AC005504
20	86.8	4.0	130281	60	AC004157
21	86.8	4.0	152209	11	HS1108D11
22	86	4.0	15421	33	PFCOMP1RA
23	85.2	3.9	14001	33	PFCOMP1RB
24	85.2	3.9	15421	33	PFCOMP1RA
					U11584 Drosophila
					U37541 Drosophila
					AL034556 Plasmodi
					AL034556 Plasmodi
					AE001398 Plasmodi
					AC005083 Homo sapi
					AL031746 Plasmodi
					AL034558 Plasmodi
					J01462 Yeast (S.ce
					AL031746 Plasmodi
					AC005504 Plasmodi
					AC004157 Plasmodi
					AC008206 Drosophi
					AL034559 Plasmodi
					AL109815 Plasmodi
					AC007465 Homo sapi
					U11584 Drosophila
					U37541 Drosophila
					AC005504 Plasmodi
					AC004157 Plasmodi
					AL034419 Human DNA
					X95275 P.falciparu
					X95276 P.falciparu
					X95275 P.falciparu


```
/db_xref="SPTREMBL:O97275"
/translaton="MLFYKICLFLSPFIIIIHCKDISSTKKLLKSGNLLNKKKRGRE
RYGIIYKSYKMTIEKRYKYNVIFKQKNNSYFSCVNYKNNNNYNTAYNIYNNP
KNKLUKYYKIVKIKKIKKIFKQKNSQNRKCKKKYFFINFTSPHDKDNKIKVLF
NIDVYENIIYSYIKSKRPPTTKIYLLGTELLSVLHMKNNVYKILLPDPNKKFK
GEIWRFLPYLIGNLQYLIMPNYUNITHSSVEISHYKKPDEFLIFEGYISNLL
ETIWMANNYNNMVKLYIINFRNFFIKDCVSKYTSRNSNNNSNNNSNNNN
HNNKNDIKKKEOYHGLVFTYIYIYWSRINEGLINCFELFFIKAEYVPPFFII
QYLLNNEFLSYEASVFSYFYIYKFKYFNFLFFTKLKVIIHIIYPLDFQMN
KLSLISLKKRNGPLFLQAKYIHNVPRIHTMKHNIISOLRKNIECVNKKYKFRV
WSNELYIILKLPKFKYVDMFIILKLSKYIKIEKVLLYICEOIEHEIKYKFR
DKLILILKRNFDNIYVNLISKSLTMKMNKNSKDKDALIYSYISKNYLTDIOI
YKNEIFNPSLKFENHANNVNLHSLSLFFYSYFPIINCCFYIYFYSFFNLIYK
FINIKNLHNSDLMFVIGLSLHHNMTPHVDONKEPKNNKNNNNHNDINR
IYKLNINCCYDHLSDNITNTYDQGYNKHIOQNHKIQOHNKHIOQNHK
IQRTGCTESNSNOOLIOIQNDKKNLITYDNSKHNLLKPCQHNIVEKQCKQ
NLINKLINTKIIIEELSKFIOEIVNLPVSLNKNIIINKYFHEFLNOEKINIRNY
INIVYNKIYLNDEERTSHGILKIKNDKDLIYHDMKELYLNMEIILYNDLLN
MKQIILGLKFNMFLOFEKILKARFICLPKKEIQKIEIMFLIYQFVRLCL
FNEUKQNNFLISODEYENIYISDKYNESAKLDNSYNSMPSNLEKNTNHHGKDNFL
DLYTHDOLFMYNKKRDRYKIIYLDNFIFNPAYVYQKDDHIDYNESVNFNDKNS
FIOUKKKINKNNNNNNNNNNIYDNTIOTVKNYSCTHNNVKIKNETNDYNPNS
TIRNQHNDVILNPNVFFYNNKLVNVDSDIFVELTCYNYLDIYKIVCLKLLK
NKKLSQSDILCIYEKLRDRYRIKLYLNKELLYLDNTYLLKVINIYKFNLY
NNISYQNLILPTINTNNNESIQILKIGLISVHKKHNLKSPFHMNNLVQNAAY
LFRNLYNQIIDLKYTEMNVYDNLTFKYKLFKNLISINVKRYQNCNSYKNNYNN
THTNNLNKQHKYIHHNNHDKGRHNNNNHNDYKDVSSSSSYIYLNKSKNIG
NIVNQLDDININKISIKYIKKDOIKDIDYMRVSKYSELMKSMKMMYDEHFNDEY
NRVQDEYDLFLYNNIOVKNICNVNTEPMANILTLNNDENILINISDDNK
KLIYKVKKELVDILNIOYVYTLKSKLDKKEYNIYLSYSHHKKKNNKILNLY
KYALLKIKKRGFNICIDADYVTKNKKSKDLSYEINKLYINLIDILKROKKNH
LHPHPTQNTKOINIKNINKLLYHONKNYKTKIIFKNYKRYKIMNLPDQRNHYH
NKRIYIKDKSLLAINKHKTNIIEKOKISTSNHLSKLKRMFSL"
complement(20528..21454)
/gene="MAL3P5.5"
CDS
complement(20528..21454)
/gene="MAL3P5.5"
/Note="predicted using hexExon: MAL3P5.5 (PFC0595C),
Serine/threonine protein phosphatase (PP2), len: 309 aa;
Similarity to serine/threonine protein phosphatases.
M.domestica serine/threonine protein phosphatase
(TR:Q42912) BLAST Score: 1005, sum P(1) = 6.9e-107; 60%
identity in 301 aa overlap."
/codon_start=1
/protein_id="CAB38970.1"
/db_xref="GI:4493934"
/db_xref="SPTREMBL:O97259"
/translaton="MAGEERKWEQLRMNPPKLLDESDRLVQRYKEILVBNVQ
SIKPPVILCDIHGCPDLLELDVGDIMNNDYIFLGDYVDRGYSNVEFYLLLLK
LFPKNITLURGNHESQITTYIGFYDECFKKYGNANAKYCTIDIFLILALDQ
IFCVHGLSPBEIKLIDQLRLJNRVOIEPHGAFGDMDSDPEDVDVWNPGRGWL
GPNTYKFNHINNLIELARAHQLAHEGYRYMFEDSTIITVMSAPNYCYRCGNAIMR
IDXYNRMQLIFKOTPDPSRNSIKKATIPVL"
25252..26157
/gene="MAL3P5.6"
join(25252..25296,25453..26157)
/gene="MAL3P5.6"
/Note="predicted using hexExon: MAL3P5.6 (PFC0600W),
Hypothetical protein, len: 250 aa"
/codon_start=1
/protein_id="CAB38972.1"
/db_xref="GI:4493936"
/db_xref="SPTREMBL:O97261"
/translaton="MKYLNKYMYIYNNYKLEKYNFLKLNKNSHMGASQNNMNV
NNYTMNELEPEKTNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
SFQHYIKLTCRADDIRNLEKRLAYERNDLNRIKIQEENKKGIGINGNFGSS
NSSSNNDNLLLRKLNRLNKNKSRKRKRKRKRKRKRKRKRKRKRKRKRKRKR
NGFKIEDYTCQNGYNDVIRSTSPNAD"
27547..28290
/gene="MAL3P5.7"
CDS
27547..28290
/gene="MAL3P5.7"
/Note="predicted using hexExon: MAL3P5.7 (PFC0605C),
```

```
Hypothetical protein, len: 248 aa"
/codon_start=1
/protein_id="CAB41709.1"
/db_xref="GI:4725991"
/db_xref="SPTREMBL:O9Y011"
/translaton="MGHGGSLNLTLPQKNNVYRKDAQYKVYHDEHKLKEEKKEIKR
KKDEFESTIILKNTKNNDSNNNNYNNFYDENGKKTNTTNCNDHNLNLFIDEKELT
AKQKKEEFLKNGHYIYDKNTQHSIYDNKNQAIISDNFKMLKLCRDMPLNKK
NKKEKTDGNAFNFFHIOKNISEHNKNTENISDLSLYCNTNNTIYTHDKKKKKOMHY
HIKKIYKQKQKKEKKRQKQKQKPK"
complement(29992..33537)
/gene="MAL3P5.8"
CDS
complement(29992..33537)
/gene="MAL3P5.8"
/Note="predicted using hexExon: MAL3P5.8 (PFC0610C),
Hypothetical protein, len: 1182 aa"
/codon_start=1
/protein_id="CAB38971.1"
/db_xref="GI:4493935"
/db_xref="SPTREMBL:O97260"
/translaton="MAHVKKKEKTEAQETPVVAKQTHAKEENNESNIATVENYIS
KNGOEIAISKNDQEIATSKNDQEIATSKNDQEIATSKNDQEIATSKNDQEIATSK
DNELKQIKEPHDIENENENRFSFTSLSSFFKEVEENSQHPFSECTHESME
DSNNVETIENAINDVLSRNSYSYSKQKNELSVTCYVCGEIVDLNINSDFIAHKL
Query Match 4.6%; Score 99.8; DB 33; Length 86829;
Best Local Similarity 45.4%; Pred. No. 0.00013;
Matches 529; Conservative 0; Mismatches 617; Indels 19; Gaps 4;
Qy 5 atagagacatgttattcaaacacataaaaatgatgctgaatgggtgattgggaacga 64
Db 40158 ATTAAAAATAATAATATGTTTCATAAAATAATAATAATAATAATAATAATA 40099
Qy 65 tatctgtttatgatgttcaggcgcaaaaatgcgactacttaataaaaatttacatt 124
Db 40098 TAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 40039
Qy 125 aaatagaaatttttttcaataaaatattattattattattattattattattatt 184
Db 40038 ATTATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 39979
Qy 185 agaaaatttgaatcccgattctccctcttcttcttcttcttcttcttcttcttca 244
Db 39978 TAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 39919
Qy 245 aaccaactctat----atgttcttcaaatgaaacttgaaacttataataataaac 300
Db 39918 TTAAATCTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 39859
Qy 301 tgaatacaatttggtatcaatcatcatcatcatcatcatcatcatcatcatcat 360
Db 39858 ATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 39799
Qy 361 ataaatcgcaaaagatttcaaatatcttccagaaaaataataataataatttgcg 420
Db 39798 ATTTATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 39739
Qy 421 tttcatggttggctgagaggatttggcaactagaaactctccacgaccattct 480
Db 39738 TATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 39679
Qy 481 ttgcaactcaaacagatggtcgagaattggtggggattttattattattcaagcata 540
Db 39678 CATTTATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 39619
Qy 541 ttcaaaaacttacttacttctcgctcgctcgctcgctcgctcgctcgctcgctc 600
Db 39618 TTATTAATTTATTAATTTATTAATTTATTAATTTATTAATTTATTAATTTATTA 39559
Qy 601 attttaccctcaaacagatttgaaggacaaaaataataataataattttcattag 660
Db 39558 TATTATATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTTATTA 39503
```



```

NDXYIENLIYSIKSKFPTTITKIYLLGTLFLLSVLIHNNKNVYKLIIFDKNKPKK
GEIWRLETPYLIIGNDIYLOILMFNLNIMSSVEISHYKLPEDFIIFTFYISNLL
FTIWNMYNENIMNVKLYTHNKNFEIKDCVSKYTSRNNNNNNINSSNNNN
HYNNKNIDKEOYHNGLYVSTYIYLYVSKYTSRNNNNNNINSSNNNN
QWILLNFESEVASIFSSYLFYVYKFKYENLRFPKTLKHVIHIVPLVDFOMN
TKLLISILAKKRGPPFLQKCYIHNPRTMKHNIISDLRKNIECVCKNVKKNV
WSNLYIITLKLFPNKEIKNVDFMILKLLSKYIKLEKVVLYICEOIEHEIYKTR
DUTLLIILKRNKNFYINISILSKMKNMSYKOLALIIYLSKNIIYDDEQI
YKNEIFNSILFENHNNINLSILFFYSYVFINCFYIYFPHFNITIK
FINIKNNULYLDMLFYIGSLIHNMYTPNHVDONKPKNNKNNHNDHN
IYIKNNINULYDMLFYIGSLIHNMYTPNHVDONKPKNNKNNHNDHN
IQORIGTHCESNNOQIIONDEKENKILITYDNNKHNLLKDPCHQNIYKDEKGO
NLKNIILNKKIEKLSFKIOETIVNLFVSLAKNIIINKYKPHFLNQEKNIRNY
INIVYNKIYLNDEEENTSHCILKNDKNDKLLDTHDMKFLINLMELIYRNDLN
MKOILLVGLFNPMFLOFEKILKRPICLPKEIKQIEEMFLYFVFVTCFL
FNEKLNKLFISQDEYENIYISDKYNESAKLDSYNMPSNLKEKTNHGGKNTL
FYLHDDLFYNNKKNRDRKTYLDNFENFYPAVYEOQKDIIDYNEVNFNNKS
FYLAKKKLNKNNNNNNNNNNIYIDTQIVTKNYSCTHNVKNETNDYVNS
TIRNOHPNDVILNPPVFNKKNLVDSIDFEYELTCVNLDIYKIYCLKLLLK
NHLKSLCSDILCIYELKINRDIYIILYKLNKELLYLDYIILYKINILVKNLY
NMISYQINKILTFINYNINISIOILLGLMILSVHKNKLSPPHMMNLVONANY
LFLKNYQIONDLKIEKMMYVNDLTFKYLFRNLIISINRYKRVONCSNYKYM
THNNLNKNEQHYTHNNDRKNNNNNNHNDYDVSSESSSYVYVYLNKSGNLG
NINVOLODDININKISYKIKKQIDIGVMYSKSELKSMKMMYDEHFNDEY
RNVCEIYEDFLYINKIOVKNINICNYTEPMALNLTNNDENILINIKNSDNK
KLIKVDKKEFLYILYNDYTYITLTKSLDKLKEYNLYSTYSNHRKKNKILNYK
KYLAKLKRGNFICIDADYIVRNKKGSKDLSYEINKLYINLLIDILKROKKNH
LHPHPTQNTKQICNINKNKLLYHONKNVKKIIFHNKRYKIKMNLDPQRNHYH
NKRIRYIKDSLLAIHNKTKNIEKOKISTSNHLSKLKRMFSL"
complement(20528. .21454)
/gene="MAL3P5.5"
complement(20528. .21454)
/gene="MAL3P5.5"
/translation="MAL3P5.5 (PFC0595c),
Serine/threonine protein phosphatase (pp2), len: 309 aa;
Similarity to serine/threonine protein phosphatases.
M.domestica serine/threonine protein phosphatase
(TR:042912) BLAST Score: 1005, sum P(1) = 6.9e-107; 60%
identity in 301 aa overlap."
/codon_start=1
/protein_id="G1:4493970.1"
/db_xref="GI:4493934"
/db_xref="SPTREMBL:O97259"
/translation="WAKEERKWIQLRNPPLKLDLRLVCRVKEILVEENNVO
SIKPPVICGDIHGOFDELDFVGGDIMNDYIFLGDYVDRGNSVETFEYLLLLK
LFPKNITLLRGHESROITTYGYDFCFKYGNAWAKYCTDIFDITLTAALYDNO
IFCVHGGISPEIKLIDQLRLINRVOEIPHEGAFGDMMSDPDEVDWVANPRGAGWF
GPNTVKFENHNNLEIARAHOLAMEGYRYMFEDSTIITVNSAPNYCYRCGVAAIMR
IDEXMNRMLIFKOTPDPSRNSIKNKATIPYL"
25252. .26157
/gene="MAL3P5.6"
Join(25252. .25296,25453. .26157)
/gene="MAL3P5.6"
/translation="MAL3P5.6 (PFC0600w),
Hypothetical protein, len: 250 aa"
/codon_start=1
/protein_id="G1:4493972.1"
/db_xref="GI:4493936"
/db_xref="SPTREMBL:O97261"
/translation="MKYLNKMYVYIYNNKLEEKYKNFLKNNHSHMGASNNVNN
NYVTWNELEEFKINNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
SFOYIKLKTCDRADDIRNLEKRLAYRONTLIRNIOEENKKGIGINGNGSES
NSSSSNDVYLLKIRNLKNTNKNRKRKRKRKRKRKRKRKRKRKRKRKRKRKR
NGFIEDYTCONGYNDFVIRSTSPNAD"
27547. .28290
/gene="MAL3P5.7"
27547. .28290
/gene="MAL3P5.7"
/translation="MAL3P5.7 (PFC0605c),
Hypothetical protein, len: 248 aa"
/codon_start=1
/protein_id="G1:4725991.1"
/db_xref="GI:4725991"

```

```

/db_xref="SPTREMBL:O97011"
/translation="MGHGGGLNLPQKKNVYRKQAYKVHYDEHKIIEEKDKBIKR
KKDEPSTLTKNNKNNEDNNNNYDENGKKTNTTNCNDHNFIDBEKELT
AKQKHEEFLIKKGHYIYDKNFNTQHNISYDKNNQAOIISDFNKKMLCERDWFNLK
NNKNETKONGANFEHLOKDNISEENKNTENISDLSLYCNTNNYITHDKKKEKKOMHY
HKKIKYQEKDKKRRQRQKRRQKPK"
complement(29992. .33537)
/gene="MAL3P5.8"
complement(29992. .33537)
/gene="MAL3P5.8"
/translation="MAL3P5.8 (PFC0610c),
Hypothetical protein, len: 1182 aa"
/codon_start=1
/protein_id="G1:4493935"
/db_xref="SPTREMBL:O97260"
/translation="NAHVKKKEKTEAOTPVVAKKEOTHAKEENNESIATVEENVIS
KNGQIAISKNDQETAIKNDQETAIKNDQETAIKNDQETAIKNDQETAIKNDQETAI
DNELKQIREFHDSIENENHNRSFSTLSSFFKEIENSVQEHFFSEGTHESME
DSNNVETIENATINDVLRNSRSTYSKOKNELTSVTCVCGETVOLNIMSDHIFAKHL"
Query Match 4.5%; Score 98.4; DB 33; Length 86829;
Best Local Similarity 45.7%; Pred No. 0.0002;
Matches 498; Conservative 0; Mismatches 581; Indels 11; Gaps 4;
QY 64 atctgtgtttattatgattcaggcgcaaaaatcgagctacttaataaaattttacatt 123
DB 39518 ATTAATATATATATTAATATATATATATATATATATATATATATATATATAT 39577
QY 124 taaattagaaattttttttatcaataaaataattatttagtttttttagaaataata 183
DB 39578 TATATATATATATATATATATATATATATATATATATATATATATATATATA 39637
QY 184 tagaaattttgaatcccgattctccctctt-cttcgctattcattcattcttaac 242
DB 39638 TAATATATATATATATATATATATATATATATATATATATATATATATATAT 39697
QY 243 caaaccaactcttatgttcttcaaat-aaacttgaaattattataataataaac 300
DB 39698 TATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAA 39757
QY 301 tcaaaccaattgggtatcattcattacatcattcattcattcattcattcattcatt 360
DB 39758 AGCAGTTATATATATATATATATATATATATATATATATATATATATATATAT 39817
QY 361 ataaactcgcaagattttcaaatatctcttcagaaaaataatacaaaattttgtcg 420
DB 39818 CTTAATATATATATATATATATATATATATATATATATATATATATATATAT 39877
QY 421 ttttcattggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 480
DB 39878 TTAATATATATATATATATATATATATATATATATATATATATATATATATAT 39937
QY 481 ttgcacttcaaacgagatggtcagaattggtggtgtgtgtgtgtgtgtgtgtgtgtgt 540
DB 39938 TAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAA 39997
QY 541 ttcaaacactctctctctctctctctctctctctctctctctctctctctctctct 600
DB 39998 ATTAATATATATATATATATATATATATATATATATATATATATATATATAT 40051
QY 601 atttttaacccttaacagtaattggaaggacaaataatttttttcaatttgatag 660
DB 40052 ATTTATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 40111
QY 661 actatttttttttttttttttttttttttttttttttttttttttttttttttttt 720
DB 40112 AGTATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 40171
QY 721 acagatgcaactaaccaataatttttttttttttttttttttttttttttttttttt 780
DB 40172 ATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 40231

```



```

QY 587 agactttcaaaatcatttttaacccttaaacagtaattgaaggacaaaataattt 646
D 7863 AAAACTATTATA-----TTTAAATATATAAAATCGTATATAATAATAATCTTTTAA 7918
QY 647 ttcaaatgtagactattttttttttgtaatttgacgaacaaacacagatttctc 706
D 7919 TTTTATTATTATAATAATTTATTATTATTATTATTATTATTATTATTATTATTATTAT 7978
QY 707 gaattttagaaacacagatgttaactaaacaaatattttttttctaaaaaataatt 766
D 7979 TTTTAAATGTTAAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 8038
QY 767 tcatgagcagatgcctcagccctgaaacaaacccctataaaaatactcacacattgacca 826
D 8039 ATAATTATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 8093
QY 827 ttgaaagtcgttcctccatcggttaaacacagatcaaacctcacacacataaacatgg- 885
D 8094 TTTTATTATTATTATTATAATAATAATAATAATAATAATAATAATAATAATAATAATA 8153
QY 886 atactccttaccacataactatttttgggttcaaatatttaattttttttaaaga 945
D 8154 AAATATTATTATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 8213
QY 946 tattaattaagaataataaagatttttttaaaaaaaagtataaaataattatttcgat 1005
D 8214 ATTAATTATTATTATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 8273
QY 1006 ttttcacatttgattgataataataatttttttttaattttttttaaataatttgc 1065
D 8274 TATTATTAAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 8333
QY 1066 aagacacttattagacatgctgttctgtttacaaaagcattcatcatttaacatt 1125
D 8334 ATATTATTATGAGTCATTATTATTATTATTATTATTATTATTATTATTATTATTATT 8393
QY 1126 aaaaaatattt 1136
D 8394 TAATAATATAT 8404

RESULT 6
AC005083/c
LOCUS Homo sapiens BAC clone CTA-281G5 from 7p15-p21, complete sequence.
DEFINITION AC005083
ACCESSION AC005083
VERSION AC005083.1 GI:4150930
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 146285)
AUTHORS Madsen,C. and Blair,T.
JOURNAL Unpublished
TITLE The sequence of Homo sapiens BAC clone CTA-281G5
REFERENCE 2 (bases 1 to 146285)
AUTHORS Waterston,R.H.
JOURNAL Direct Submission
TITLE Submitted (12-JUN-1998) Genome Sequencing Center, Washington
MO 63108, USA
REFERENCE 3 (bases 1 to 146285)
AUTHORS Waterston,R.
JOURNAL Direct Submission
TITLE Submitted (12-JAN-1999) Department of Genetics, Washington
UNIVERSITY, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE 4 (bases 1 to 146285)
AUTHORS Waterston,R.
JOURNAL Direct Submission
TITLE Submitted (21-DEC-1999) Department of Genetics, Washington
UNIVERSITY, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT On Jan 12, 1999 this sequence version replaced gi:3212908.

```

----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu/gsc>
Contact: sapiens@wustl.wustl.edu
----- Summary Statistics

Center project name: H_RG281G05

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double-stranded or sequenced with an alternate chemistry; an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgrl.nih.gov/DIR/FTP/CHR7>, send mailto:egreen@nhgri.nih.gov, or see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

Clone CTA-281G5 is from a release of the human BAC library CITB-HS-A. The library contains cloned DNA from human sperm. See: Shizuya et al., Proc. Natl. Acad. Sci. USA 89:8794-7 (1992); U-J. Kim et al., Genomics 34:213-8 (1996). The clone is available from Research Genetics, Inc. (<http://www.resgen.com>).
VECTOR: pBeloBAC11
Selection: chloramphenicol

NEIGHBORING SEQUENCE INFORMATION:

Actual start of this clone is at base position 1 of CTA-281G5;
actual end is at 146285 of CTA-281G5

BAC CTA-281G5 contains an E. coli transposon from 10982 to 12345 that is not represented in the submitted sequence.

FEATURES	Location/Qualifiers
Source	1..146285 /organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="7" /map="7p15-p21" /clone="CTA-281G5" /clone_lib="CITB-HS-A" 322..356 /rpt_family="(TAAAA)n" 808..952 /rpt_family="MER1_type" 1172..1194 /rpt_family="AT_rich" 1264..1720 /rpt_family="L1" 1848..1881 /rpt_family="AT_rich" 1882..2171 /rpt_family="Alu" 3782..3804 /rpt_family="AT_rich" 4794..4826 /rpt_family="AT_rich" 5764..6034 /rpt_family="purine-rich" 6891..7126


```
repeat_region /rpt_family="L1"
7133..7263
repeat_region /rpt_family="L1"
7455..7481
repeat_region /rpt_family="AT_rich"
7504..7554
repeat_region /rpt_family="AT_rich"
8320..8348
repeat_region /rpt_family="(CA)n"
8466..8533
repeat_region /rpt_family="AT_rich"
9611..9733
repeat_region /rpt_family="MIR"
9946..10172
repeat_region /rpt_family="MER1_type"
10437..10477
repeat_region /rpt_family="(CA)n"
10581..10681
repeat_region /rpt_family="(TAAA)n"
10759..11049
repeat_region /rpt_family="AT_rich"
11050..11235
repeat_region /rpt_family="(CAATA)n"
11408..11526
repeat_region /rpt_family="(CATAT)n"
11768..11872
repeat_region /rpt_family="(CATTA)n"
12002..12110
repeat_region /rpt_family="(CATAT)n"
12057..12168
repeat_region /rpt_family="(TA)n"
12381..12490
repeat_region /rpt_family="(CAATA)n"
12562..12680
repeat_region /rpt_family="(CTATA)n"
13131..13367
repeat_region /rpt_family="L2"
13378..13540
repeat_region /rpt_family="MER1_type"
13557..13744
repeat_region /rpt_family="L2"
13745..14592
repeat_region /rpt_family="L1"
14593..14616
repeat_region /rpt_family="L2"
16078..16548
repeat_region /rpt_family="L1"
18492..18692
repeat_region /rpt_family="MER1_type"
19334..19422
repeat_region /rpt_family="L1R41"
19880..20137
repeat_region /rpt_family="L1R41"
20504..20541
repeat_region /rpt_family="MER1_type"
21092..21607
repeat_region /rpt_family="MaLR"
22221..22499
repeat_region /rpt_family="Alu"
22500..22523
repeat_region /rpt_family="(CAAAA)n"
22535..22857
repeat_region /rpt_family="MER4-group?"
23203..23380
repeat_region /rpt_family="L1"
23381..23643
repeat_region /rpt_family="Alu"
23644..24770
repeat_region /rpt_family="L1"
25826..25846
repeat_region /rpt_family="AT_rich"
26130..26254
repeat_region /rpt_family="MIR"
```

```
repeat_region 26529..26563
/rpt_family="AT_rich"
repeat_region 26792..27116
/rpt_family="Alu"
repeat_region 30744..30982
/rpt_family="MIR"
repeat_region 31032..31739
/rpt_family="L1"
repeat_region 32332..32448
/rpt_family="MIR"
repeat_region 32536..32730
/rpt_family="MIR"
repeat_region 33282..33333
/rpt_family="AT_rich"
repeat_region 35261..35288
/rpt_family="(CATTT)n"
repeat_region 35571..35598
/rpt_family="(CA)n"
repeat_region 35924..36007
/rpt_family="L2"
repeat_region 36476..37873
/rpt_family="L1"
repeat_region 38324..38899
/rpt_family="Retroviral"

Query Match 4.3%; Score 92.6; DB 39; Length 146285;
Best Local Similarity 44.5%; Pred. No. 0.00089;
Matches: 467; Conservative 0; Mismatches 569; Indels 13; Gaps 2;

QY 109 ataaattttacatttaaaatttagaattttttttatcatcaataaataattattattattttt 168
DB 11581 ATATAATATATATTTATATATATATATATATATATATATATATATATATATATATATATG 11522
QY 169 attagaaaatttaattagaattttgaattcccgatttctctctctctctctctctctctct 228
DB 11521 ATATGTAATATATATATATATATATATATATATATATATATATATATATATATATATAT 11462
QY 229 catcatttttcaaccacaaactcttatgtcttccaaattagaaacttagaaattatttaa 288
DB 11461 ATTAATATGATATATATATATATATATATATATATATATATATATATATATATATATAT 11402
QY 289 ttataattaaactgaaacaaactttggtatcaattcatcacatgcttagtaataaaatgc 348
DB 11401 ATTATTCATATATCTAAATATATATATATATATATATATATATATATATATATATATATAT 11342
QY 349 gataaattgataaaactgcgaaagattttacaaatattttcacaatacttttcagaaaaactaa 408
DB 11341 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 11282
QY 409 caaattttgtcttttcattggtggtggtggtggtggtggtggtggtggtggtggtggtggt 468
DB 11281 ATCATTTATATATTTCTATATATATATATATATATATATATATATATATATATATATATAT 11222
QY 469 acggaccattctttgcacttcaactaaacgagatggtcagaattggtggtggtggtggtggtggt 528
DB 11221 TATAAATATATCTTTTATACATAAATATATATATATATATATATATATATATATATATATAT 11162
QY 529 aagcatacccttttcaaaacttctacttacttctcgctcgctcgctcgctcgctcgctcgctcg 588
DB 11161 TAATATTTATATATATATATATATATATATATATATATATATATATATATATATATATAT 11102
QY 589 actttcaaaatcatttttaacccctaaacagtaaatgtgaaggacaaaaaataattttt 648
DB 11101 ATTTATATACAAATATATATATATATATATATATATATATATATATATATATATATATAT 11050
QY 649 caaatttgatagactatttttttttttttttttttttttttttttttttttttttttttttt 708
DB 11049 TTTATATATATACAAATATATATATATATATATATATATATATATATATATATATATATAT 10990
QY 709 attttaggaaccacagatgtaactaaacaaatattttatttttttttttttttttttttttttt 768
DB 10989 AATGTATCATTTATATATATATATATATATATATATATATATATATATATATATATATAT 10930
```


DEFINITION Yeast (S.cerevisiae) mitochondrial Ser-tRNA and ATPase proteolipid genes.

ACCESSION J01462 J01463

VERSION J01462.1 GI:343938

KEYWORDS ATPase; proteolipid; transfer RNA; transfer RNA-Ser.

SOURCE Yeast (S.cerevisiae) mitochondrial DNA; cytoplasmic petite mutants (rho-) ds400/a3 [1], ds401 [2].

ORGANISM Mitochondrion Saccharomyces cerevisiae

REFERENCE Eukaryota; Fungi; Ascomycota; Saccharomycetales; Saccharomycetaceae; Saccharomyces.

AUTHORS 1 (bases 1026 to 2806)

TITLE Macino,G. and Tzagoloff,A. Assembly of the mitochondrial membrane system: The DNA sequence of a mitochondrial ATPase gene in Saccharomyces cerevisiae

JOURNAL J. Biol. Chem. 254, 4617-4623 (1979)

MEDLINE 79173209

REFERENCE 2 (bases 1 to 5371)

AUTHORS Tzagoloff,A., Nobrega,M., Akai,A. and Macino,G.

TITLE Assembly of the mitochondrial membrane system. Organization of yeast mitochondrial DNA in the OLI1 region

JOURNAL Curr. Genet. 2, 149-157 (1980)

COMMENT [2] notes that even though there are other possible coding sequences in the ds401 segment, none are sufficiently long to code for a gene product of the size of the var1 protein. a model for the mitochondrial RNA processing is proposed.

FEATURES

source Location/Qualifiers

1..5371

/organism="Saccharomyces cerevisiae"

/organelle="mitochondrion"

/db_xref="taxon:4932"

1..5371

/gene="OLI1"

/db_xref="SGD:S0007274"

1372..1602

/gene="OLI1"

/note="ATPase proteolipid"

/codon_start=1

/db_xref="SGD:S0007274"

/transl_table=3

/protein_id="AAA32169.1"

/db_xref="GI:343938"

/translation="MQLVLAAYKIGAGISTIGLLGAGIGIAIVFAALINGVSRNPSIKDTVPFMAIFGALSEATGLFCLMVSFLLFGV"

2856..2942

/gene="OLI1"

/note="codon recognized: UCA; Ser-tRNA"

/product="tRNA-Ser"

/db_xref="SGD:S0007274"

2294 a 352 c 368 g 2191 t 166 others

BASE COUNT 82 bp 5' to SstII site, at wild-type map units 79.1.

ORIGIN

Query Match 4.2%; Score 91; DB 7; Length 5371;

Best Local Similarity 45.7%; Pred. NO. 0.0039;

Matches 478; Conservative 0; Mismatches 550; Indels. 17; Gaps 4;

QY 16 ttattcaaaacataaataatgatgctcaaaattggtggtggaacgatactctgtttat 75

DB 4287 TTATCTATATCAATTTAACTAATATATATTTACTTAAATAATATATATATAT 4228

QY 76 tatgatttcaggcgcaaaatgcgacttaataaaattttacatttaaaattagaatt 135

DB 4227 GATATTTAATAGGTTCAATAGTTACTTTTATTAATAATACTAAATATATACAT 4168

QY 136 ttttttcaataaataattattattgatttttagaataatttaataaattttg 195

DB 4167 AAATTATTATTATATTATTATATATATTATTTATTTATTTATTTATATATAA 4108

QY 196 aatccccgattctcctctcttcttcgctatcatcttcttaacaaacccaatctta 255

DB 4107 TAATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATAA 4048

QY 256 tatgttcttcaattagaacttgaattattattataattataaactgaacaaatttgg 315

DB 4047 TATTTAAATTTATTTAAACGTGTTTAAATTAAGTTTACTTATAAATTTTATTAATAT 3988

QY 316 atcaattcat--atacatgcttagtaataaaatcgataataatgataaaatctgcaaa 373

DB 3987 TTATATTATTTAATTTAAAGTTATTTTATATATAATTAATTAATTAATTTATCAT 3928

QY 374 agattttacaatatcttttcagaaaaaattataa--caaatgttgcgttttcattggtg 431

DB 3927 TTATAGTATTTTAAATTTCTATATTTTATTAATAATAAATTTGTAGATTTCAATTTAT 3868

QY 432 ttgggtcagaggaggttggcactatagaaactcctcagcgaccattcttctgaactcaa 491

DB 3867 TTATATTATTTAAATTTCTTAAATTTTATTAATTAATTTCCCGGGACCAATCCGGTGAA 3808

QY 492 ctcaacgcatggtcagaattggtgggattttattattcaagcatatccctttcaaaacttc 551

DB 3807 CAACCGGATGCGCGCGGGTATTATATTATTATTATTATTCTTTTATATTAT 3748

QY 552 ctacttacttcgctggttcggttaactgtaacttagacttcacaaatcatttttaacccc 611

DB 3747 TTATTTTATTTATTTATTTAT---ATCTAATAATATATTTTAAATAAATTTCTATT 3692

QY 612 ctaaaacagtaatttgaaggacaaaataatttttcaaaatttgcatactatttttt 671

DB 3691 TTATGAAATCTGATAGATTATTAAATAAATATTATTATTATTAACTTTTATTATTA 3632

QY 672 ttgttaatttgacgaacaaaa-----ccagatttctcctgaatttttaggaaccac 722

DB 3631 TTGATAAAATTTATTTAATAATTTTAAATTTTCTATTTTATTTTACTTTTATCTA 3572

QY 723 agatgtaactaaacaaatatttattttctcaaaacaaatttcacgagcagtcgt 782

DB 3571 TTATATATATATATTAATTAATTAATTAATAATAAGATAATATATATATAATA 3512

QY 783 cagcccatgaaaaaaccttataaaatctcacattgacattgaaagtcgttct 842

DB 3511 TGCATCCCTTATTATATTATTATTATTAAATAATAATATTAAAGTATTATTG 3452

QY 843 cccatgggttaaccagatcaaacctcacatcccaacatacattggtatctccttacaatc 902

DB 3451 CAGTTATTATTATACATATATTTTATTAATAATAATAATAATTAATAATTAATA 3392

QY 903 atactaaatttttgggttaaatataatcatcatttttttaagataatttaagaataa 962

DB 3391 TTATAAATTTATATATGATAATTTATAAAATAAATATTATAATTTATATAAATAA 3332

QY 963 aaagatttttcaaaaaatgataaaattattattcattgatttttcatactattgatt 1022

DB 3331 AAATATTATAAATAAATAATATATATATTTATTAAATAATATATATAATAATAAAT 3272

QY 1023 ttgataataaataatatttttttttaa 1047

DB 3271 ATATATATATATATAATAATAATAA 3247

RESULT 10

PFMALIP3 67970 bp DNA -INV 15-DEC-1999

LOCUS Plasmodium falciparum MALLIP3, complete sequence.

DEFINITION AL031746

ACCESSION AL031746

VERSION AL031746.9 GI:6594243

KEYWORDS HTG.

SOURCE malaria parasite P. falciparum.

ORGANISM Plasmodium falciparum

REFERENCE Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.

AUTHORS 1 (bases 1 to 67970)

Bowman,S., Churcher,C., Harris,B., Harris,D., Lawson,D., Quail,M. and Barrell,B.

Direct Submission

Submitted (24-SEP-1998) P.falciparum Genome Sequencing Consortium, The Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK


```

IANEVKNHSTNIENIFSLINDSKSPVLTKFLKEFGSIFPHMLNNVPKLFLDLQ
RNPILHILGLIVILAAIYVFENKNEC
36744..36749
/genes="MAL1P3.05"
/notes="potential splice donor sequence, aag/gtatga"
36854..36863
/genes="MAL1P3.05"
/notes="potential splice acceptor sequence"
complement(38049..40284)
/genes="garp"
complement(join(38049..39995,40210..40284))
/genes="garp"
/notes="MAL1P3.06, gap, len: 673 aa, similarity: almost
identical to GARP_PLAFF (678 aa), fasta scores: 97.6%
identity in 678 aa overlap"
/codon_start=1
/product="hypothetical gap protein"
/protein_id="CAB63561.1"
/db_xref="GI:6594249"
/translation="MNVLFSYNICILFFVCTLNFSTKCFSGNLLKNONINKSFD
ITGRLLNETELKNDKSDSETLKEEDKDDVPTTSDNLKNAHNNIEISSTD
TNIINVNDKNSVLYTGOKHKKKPKNATEHGENLYEEMVSEINNAOGGLLSPOYRE
DKKKEENSVMISYTGOKHKKKPKNATEHGENLYEEMVSEINNAOGGLLSPOYRE
QCGGLISVHETSNDTKNDKENISEDKEDHQEEMKLTLDKDKERKQEKEMKEQE
KLEKKKQOEERKQEKERKQEKERKQEKERKQEKERKQEKERKQEKERKQEK
HOKNEETMOODPQSEETNNEIMVPLSPDTVTPEHKEGHEKEHEKEHEKEG
EKHEEHEKEHEKEHEKEHEKEHEKEHEKEHEKEHEKEHEKEHEKEHEKEHE
DKDVEIINLEDEKAECEOHITVESRPLSOPCKLIDPEQLTLMDSKVEKNLSIO
EOLCTIGRVNVVRDNNHKKMAIEAELOKHVDKEEDKKESEKVESEKEVO
EGBEVEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDE
DDADEDDDDADEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDE
complement(39996..40005)
/genes="garp"
/notes="potential splice acceptor sequence"
complement(40204..40209)
/genes="garp"
/notes="potential splice donor sequence, aag/gtaaca"
45401..50233
/genes="MAL1P3.07"

Query Match 4.1%; Score 89.6; DB 33; Length 67970;
Best Local Similarity 47.3%; Pred. No. 0.0027;
Matches 506; Conservative 0; Mismatches 544; Indels 20; Gaps 7;

QY 93 aaatgcgacttaataaaatttacatttaataattttttttatcaataaata 152
Db 10119 AAATTTAATATGAAAAATAATTTTATTTTAAATAATATATTTAATAATA 10060
QY 153 ttaatttatttagtttattagaataatttagaataattttgaatccccgattctct 212
Db 10059 TAAATAATTAATTTAATAAGTAAAAATATATTTTATTTTAAATAA---ATATATAAT 10003
QY 213 cctttcttcgctatcatcatcttcttaacaaacaaatcttatgtcttccaaatag 272
Db 10002 ATTTAATAATAATTTAATAATTTAATAAGAAATTTAATAATATATTAATAATTTA 9943
QY 273 aactgaaattattaataataactgaacaaatatttggtatcaatcatcatacatg 332
Db 9942 TATTAAATAATTAATTTAGTTTAAATAAGTATATTAATAA---AATAATAATTTT 9885
QY 333 cttagtaataaaatgcgataaataattgataaactgcgaagaattttcaaatatcttt 392
Db 9884 ATTATTATAATAAATAATAATATT-----ATTATAATAATATTAAATAATATT 9830
QY 393 cagaaaaataataacaaatttgcgttctggtgtgtgctgagagattggc 452
Db 9829 ATTTTAAATAATAATTTAATAATAAATGGAATAATAATAATAATGATATTAATACATT 9770
QY 453 actagaactctctcagcaccattctttgcacttcaactcaacgagtgacgaattgg 512
Db 9769 AATAAATAAATAATTTTAAAGTTAATTTAATAATATATTTATTTAATAATATTATA 9710
QY 513 tg9ggattttatcaagcatatcccttcaaaacttctacttacttctgctgctcg 572
```

```

Db 9709 T--ATAATTATATAAAATATATATATATATAAAATAAAATTTAATATTATAAT 9652
QY 573 taatcgtgaacattagacattcaaaatcatttttaaccctcaacagtaaatgtgaagg 632
Db 9651 TTACATATATAAATATATATATATATATATATATATATATATATATATATTAAT 9592
QY 633 caaaaaataatttttcaaatgtgagactatttttttttttttttttttttttttttttt 692
Db 9591 TAATTAATAATATATATATATATATATATATATATATATATATATATATATATAT 9537
QY 693 accagatttattcctggaatttttaggaacacagatgtaactaaacaaatatttttttttt 752
Db 9536 ATTAAATATATATTTAAATATATTAAGTTATTAATTAATAATAATAATAATAATAATG 9477
QY 753 tctaaaaacaaatttccatggcagcatgctccagccatgaaacaaacccctataaaatat 812
Db 9476 TGTAAATTTATTTTAAATAATATATATATATATATATATATATATATATATATAT 9419
QY 813 ctacacattgaccattgaaaagttctgtctcccatgggtaaccagatcaaacatcatcc 872
Db 9418 TTATATATATAAAATATATAAATAATATATATTTCTTTTATTTTAAATAATATATTAAT 9359
QY 873 aaacataacatggatctctccctaccacatcataactatttttgggttaaatataatc 932
Db 9358 ATTATATATATATATTTTATTTATATATATGTTACTAAAAATAATAATAATAATAATAT 9299
QY 933 attatttttaagatatataataaagaatttaaaagatttttttaaaaaaatgataaatta 992
Db 9298 ATTAATAATATATAAATAATAAATAATAATAATAATAATAATAATAATAATAATAATAAT 9239
QY 993 tattattcatgat-tttttacacatttgatttttgataataataataataatttttttaattc 1051
Db 9238 TATTTTATATATATTTTATATATATTTAATTTAATTTTATAGCGTATATATTTATTTTATAT 9179
QY 1052 ttaaaaaatgtgcaagacattattagacatagctgttctgtgttttacaacaaagattca 1111
Db 9178 ATTAAATAATTTTAAATTTTAAATTTTAAATTTTAAATTTTATATATATTTAATAATTTA 9119
QY 1112 tcaattcaacataaaaaattttaatacaacagtagaattcttctgt 1161
Db 9118 ATATAATAAATGTTTATTTTATTTAATAATTTATATATATATATATATTTT 9069

RESULT 11
AC005504/c
LOCUS
DEFINITION
AC005504
Plasmodium falciparum chromosome 12, *** SEQUENCING IN PROGRESS
***, 3 unordered pieces.
AC005504
HTG: HTGS, PHASE1
Plasmodium falciparum
malaria parasite p. falciparum.
ORGANISM
Plasmodium falciparum
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE
1 (bases 1 to 104992)
Hyman, R.W., Qin, F., Fung, E.L., Qin, F., Tamaki, T., Kurdi, O.B., Conway, A.B.
and Davis, R.W.
Plasmodium falciparum 3D7 chromosome 12
TITLE
JOURNAL
REFERENCE
2 (bases 1 to 104992)
Hyman, R.W., Qin, F., Fung, E.L., Conway, A.B. and Davis, R.W.
Direct Submission
Submitted (21-AUG-1998)
Stanford DNA Sequencing and Technology
Center, Stanford University, 855 California Avenue, Palo Alto, CA
94304, USA
COMMENT
On Apr 2, 1999 this sequence version replaced gi:4337172.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
```


Qy 803 ataaaaatatcacattgaccattgaaaagttcgttctcccatgggtaaccagatcaa 862

Matches

COMMENT On Mar 15, 2000 this sequence version replaced qi:6652498.

cn 4.18; score 89.6; DB 60; Length
1 similarity 47.38; Pred. No.: 0.0022;

Query Match 4.1%; Score 89.6; DB 60; Length 130281;
Best Local Similarity 47.3%; Pred. No. 0.0022;
Matches 534; Conservative 0; Mismatches 574; Indels 20; Gaps 8;

Db 52687 TTTATTCAATTTTGAATGTCATATTTTATATTTTCAAGAAAAAATAAATATATAT 52746
Qy 884 ggatactcccttaccactactaattatttgggttaaatatttaattcatttttttaa 943
Db 52747 ATATATATATATATTTTATATATGTTTTTATTTTAAAGTTTAAAGTTTTTATAA 52806
Qy 944 gatattaattaagaatttaaaagattttttaaaaa---atgtataaaattatattattc 1000
Db 52807 ATTCCAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 52866
Qy 1001 atgatttttcatacatcttgatttttgataataatatttttttgaatttccttaaaaaa 1060
Db 52867 TATGTTTTCTCTAAATTGAAAATTGAATAAGATTATTTATTTTATTTTATTTTATTTT 52926
Qy 1061 gtgcaagacacattattagacatagtctgtctgtttacaaaagcattcatcattta 1120
Db 52927 ATTAACAT 52986
Qy 1121 acattaaaaaatatt 1135
Db 52987 ATATTAAATTATTATT 53001

Search completed: September 2, 2000, 02:38:02
Job time: 11146 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 2, 2000, 00:25:01 ; Search time 2111.72 Seconds
(without alignments)
3286.927 Million cell updates/sec

Title: US-09-464-528-15
Perfect score: 1574
Sequence: 1 atcgtatagacatgtatt.....gttttgaagtataaagatg 1574

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 5247842 seqs, 2204914090 residues

Total number of hits satisfying chosen parameters: 10495684

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*

1: em_est1:*
2: em_est2:*
3: em_est3:*
4: em_est4:*
5: em_est5:*
6: em_est6:*
7: em_est7:*
8: em_est8:*
9: em_est9:*
10: em_est10:*
11: em_est11:*
12: em_est12:*
13: em_est13:*
14: em_est14:*
15: em_est15:*
16: em_est16:*
17: em_est17:*
18: em_est18:*
19: em_est19:*
20: gb_est1:*
21: gb_est2:*
22: gb_est3:*
23: gb_est4:*
24: gb_est5:*
25: gb_est6:*
26: gb_est7:*
27: gb_est8:*
28: gb_est9:*
29: gb_est10:*
30: gb_est11:*
31: gb_est12:*
32: gb_est13:*
33: gb_est14:*
34: gb_est15:*
35: gb_est16:*
36: gb_est17:*
37: gb_est18:*
38: gb_est19:*
39: gb_est20:*
40: gb_est21:*
41: gb_est22:*
42: gb_est23:*
43: gb_est24:*

44: gb_est25:*
45: gb_est26:*
46: gb_est27:*
47: gb_est28:*
48: gb_est29:*
49: gb_est30:*
50: gb_est31:*
51: gb_est32:*
52: em_est20:*
53: em_est21:*
54: em_est22:*
55: em_est23:*
56: em_est24:*
57: em_est25:*
58: em_est26:*
59: gb_est33:*
60: gb_est34:*
61: gb_est35:*
62: gb_est36:*
63: gb_est37:*
64: gb_est38:*
65: em_est27:*
66: em_est28:*
67: em_est29:*
68: em_est30:*
69: gb_est39:*
70: gb_est40:*
71: gb_est41:*
72: gb_est42:*
73: gb_est43:*
74: gb_est44:*
75: em_est31:*
76: em_est32:*
77: em_est33:*
78: em_est34:*
79: gb_est45:*
80: gb_est46:*
81: gb_est47:*
82: em_est35:*
83: em_est36:*
84: em_est37:*
85: gb_est48:*
86: gb_est49:*
87: gb_est50:*
88: gb_est51:*
89: gb_est52:*
90: gb_est53:*
91: gb_est54:*
92: gb_est55:*
93: gb_gss1:*
94: gb_gss2:*
95: gb_gss3:*
96: gb_gss4:*
97: em_gss1:*
98: em_gss2:*
99: em_gss3:*
100: em_gss4:*
101: gb_gss5:*
102: gb_gss6:*
103: gb_gss7:*
104: gb_gss8:*
105: gb_gss9:*
106: em_gss5:*
107: em_gss6:*
108: em_gss7:*
109: em_gss8:*
110: em_gss9:*
111: em_gss10:*
112: em_gss11:*
113: gb_gss10:*
114: gb_gss11:*
115: em_gss12:*
116: gb_gss12:*

117: gb-gss13:.*
 118: gb-gss14:.*
 119: gb-gss15:.*
 120: gb-gss16:.*
 121: gb-gss17:.*
 122: gb-gss18:.*
 123: gb-gss19:.*
 124: em-gss13:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	95	6.0	1101	122	CNS00EVL	AL069706 Drosophila
C 2	92	5.8	1101	122	CNS00EVL	AL069706 Drosophila
C 3	89.2	5.7	1101	122	CNS00EVL	AL069440 Drosophila
C 4	88.6	5.6	928	122	CNS00DKY	AL071865 Drosophila
C 5	87	5.5	987	123	CNS014PQ	AL104456 Drosophila
C 6	85.2	5.4	1101	122	CNS003BD	AL064091 Drosophila
C 7	84	5.3	1101	122	CNS0039G	AL063921 Drosophila
C 8	83.8	5.3	843	122	CNS00CS1	AL059666 Drosophila
C 9	83	5.3	1101	122	CNS0021J	AL061936 Drosophila
C 10	82.8	5.3	1101	123	CNS0161I	AL106896 Drosophila
C 11	81.6	5.2	1225	123	CNS0161D	AL106171 Drosophila
C 12	81.2	5.2	928	122	CNS00DKY	AL071865 Drosophila
C 13	81.2	5.2	1201	123	CNS0161I	AL106627 Drosophila
C 14	81	5.1	1043	123	CNS0145P	AL103735 Drosophila
C 15	80	5.1	1200	123	CNS016CO	AL106578 Drosophila
C 16	79.8	5.1	996	122	CNS000FH	AL071063 Drosophila
C 17	79.2	5.0	1101	122	CNS001FB	AL060732 Drosophila
C 18	78.6	5.0	1101	122	CNS00YWL	AL059627 Drosophila
C 19	77.4	4.9	1101	122	CNS000B8	AL063632 Drosophila
C 20	77.4	4.9	1101	122	CNS000B8	AL063632 Drosophila
C 21	76.8	4.9	1101	122	CNS000FYG	AL069440 Drosophila
C 22	76.8	4.9	1201	123	CNS0167M	AL071206 Drosophila
C 23	76	4.8	1101	122	CNS00B01	AL057419 Drosophila
C 24	75.6	4.8	734	122	CNS010MP	AL059163 Drosophila
C 25	75.6	4.8	1101	122	CNS0021J	AL061936 Drosophila
C 26	75.6	4.8	1101	122	CNS0039G	AL063921 Drosophila
C 27	75	4.8	1201	123	CNS0167M	AL106396 Drosophila
C 28	74.8	4.8	1101	122	CNS001VP	AL078809 Drosophila
C 29	74.4	4.7	987	123	CNS014PQ	AL104456 Drosophila
C 30	74.4	4.7	996	122	CNS000FH	AL071063 Drosophila
C 31	74.4	4.7	1101	122	CNS003DQ	AL064580 Drosophila
C 32	74.4	4.7	1101	122	CNS000FYG	AL071206 Drosophila
C 33	74.2	4.7	1043	123	CNS0145P	AL103735 Drosophila
C 34	74.2	4.7	1225	123	CNS0161D	AL106171 Drosophila
C 35	73.2	4.7	994	123	CNS015XG	AL106030 Drosophila
C 36	73.2	4.7	1101	122	CNS001T2	AL078714 Drosophila
C 37	73	4.6	1101	122	CNS003BD	AL064091 Drosophila
C 38	72.8	4.6	820	120	B11728	B11728 T8D18-Sp6.1
C 39	72.6	4.6	860	123	CNS018BP	AL109135 Drosophila
C 40	72.2	4.6	1101	122	CNS00EPO	AL069493 Drosophila
C 41	72	4.6	1101	122	CNS000PMC	AL070972 Drosophila
C 42	71.8	4.6	843	122	CNS00CS1	AL059666 Drosophila
C 43	71.6	4.5	1101	122	CNS0022H	AL097139 Drosophila
C 44	71.6	4.5	1248	120	B11336	B11336 F19M10-Sp6
C 45	71.4	4.5	770	113	AQ740708	AQ740708 HS_5507_A

ALIGNMENTS

RESULT 1
 CNS00EVL/c 1101 bp DNA GSS 04-JUN-1999
 LOCUS Drosophila melanogaster genome survey sequence T7 end of BAC:
 DEFINITION BACR29823 of RPCI-98 library from Drosophila melanogaster (fruit

fly), genomic survey sequence.

AL069706

AL069706.1 GI:4949849

KEYWORDS

GSS

SOURCE

fruit fly.

ORGANISM

Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;

Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

Genoscope.

1 (bases 1 to 1101)

Direct Submission

Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :

BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr

- Web : www.genoscope.cns.fr)

Determination of this BAC-end sequence was carried out as part of a

collaboration with the Berkeley Drosophila Genome Project (BDGP).

The BDGP is constructing a physical map of the Drosophila

melanogaster genome using these BACs. For further information

please see <http://www.fruitfly.org> The BDGP Drosophila

melanogaster BAC library was prepared by Kazutoyo Osoegawa and

Aaron Mammosier in Pieter de Jong's laboratory in the Department of

Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,

NY. The library is named RPCI-98 and was constructed by partial

ECORI digestion of Drosophila DNA provided by the BDGP from the

isogenic strain y2; cn bw sp, the same strain used for the BDGP's

P1 and EST libraries. A more detailed description of the library

and how to order individual BAC clones, the entire library, or

filters for hybridization from the BACPAC Resource Center can be

found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

Location/Qualifiers

1. 1101

/organism="Drosophila melanogaster"

/db_xref="taxon:7227"

/clone.lib="RPCI-98"

/clone="BACR29823"

/note="end : T7"

BASE COUNT 419 a 91 c 60 g 299 t 232 others

ORIGIN

Query Match 6.08; Score 95; DB 122; Length 1101;

Best Local Similarity 34.78; Pred. No. 8.3e-07;

Matches 197; Conservative 100; Mismatches 270; Indels 0; Gaps 0;

QY 573 taatcggtacattagacatttcacaaatcatttttaacccttaacacagtaatttgaaagga 632

Db 1059 TWTWTAT 1000

QY 633 caaaaataatattttcacaatttgatagactattttttttgttaattgcgaacacaa 692

Db 999 WTAWAT 940

QY 693 accgattatctcgaatttttaggaacacacagatgtaactaaacacattattatttt 752

Db 939 AWWTTTATTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 880

QY 753 tctaaacacaaatttcattcgccagcagcctccagccctgaaacacacacacacacacacac 812

Db 879 TTTTWTAT 820

QY 813 ctacacattgaccattgaaaggttcgtctcccatgggtaaccagatcaaacacacacac 872

Db 819 WATWTAT 760

QY 873 aaacataacatggatattcctcccaatcataactaactatttttgggttaaatattatc 932

Db 759 WAT 700

QY 933 attatttttaagatataacaaagataaaagatttttttaaaaaaatgataaataa 992

Db 699 TTTTWTAT 640

[illegible]

BASE COUNT	366 a	66 c	104 g	351 t	214 others
ORIGIN					

Query Match 5.7%; Score 89.2; DB 122; Length 1101;
Best Local Similarity 40.2%; Pred. No. 7.5e-06;
Matches 235; Conservative 74; Mismatches 264; Indels 12; Gaps 2;

Qy 540 ttccaacttcctacttactctggtgcgttaatcggtaacattagactttcaaat 599
||| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db 999 TTTTMAATATYCWATAATTTTMAATACAWWYATTTTWTATACATAA'TTTTTTATAT 940

Qy 600 catttttaacccctaaacagtaaatgtgaagagcaaaaaataatatttttcacaatttgata 659
 . |||:| ||| : ||: | : | : | : | : ||||| ||
 Db 939 ACAATTTWAAAAATAAAAAWTAACWAAATTTAWAAAHCAATTTWTTTAAWWTAAATTTWATT 880

[illegible]

Oy 720 cacagatgtaactaaaccaaatttttctataattttcctaaaacaaaaatttcattggcgagcatg 779
| : | | : | : |||| : ||| : |::| : |
Ox 819 AAGAAATTATGAGAATTCTTMMATAATACACTCAATTAATTTAATGAAGCCTGCGCATATAT 760

Qy 780 cctcagcccatgaaaaaaaccttataaaaaatatctacacattgaccattgaaaagttcgt 839

Qy 840 tctcccatgggtaaccagatcaaaactcacatccaaacataacatggatattctcttacc. 899

900 atcatactaattatttttggggttaataattattatcattattttttaagatatatttaataagaaa 959

Qy 960 ttaaagatttttaaaaaaatgtataaaaattatatattcatgattttttcatacatgtg 101

Qy 1020 attttgataataatataattttttaaattcttaaaaaattgttgcaagacactattag 107

D6 526 ATTTT-----TATTATTATTTTAAAAAATTTTATTATATATTAATTAATAATTTAAITTA 47

Qy 1080 acatagtcttgctgtttacaaaagcattcatcataatacat 1124
 : | | | | : : : | | | | : : : | |

471 TYTTTTTTTATATCTTTCMAAWPATTTTTCCTCCCYTTTWTGAT 427

CNS00DKY	Locus	CNS00DKY	928 bp	DNA	GSS	04-JUN-1999
	DEFINITION	Drosophila melanogaster genome survey sequence T7 end of BAC #				

fly), genomic survey sequence.
 AL071865
 AL071865.1 GI:4948170

fruit fly, *Drosophila melanogaster*
ORGANISM
SOURCE
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

REFERENCE
AUTHORS
1 (bases 1 to 928)
Genoscope.

JOURNAL
TITLE
Direct submission
Submitted (02-JUN-1999) Genoscope - Centre National de Séquençage
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cnr.fr)
- Web : www.genoscope.cnr.fr

COMMENT
Determination of this BAC-end sequence was carried out as part of collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome utilizing these BACs. For further information

please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Kammer in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPc1-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y¹; cn bw sp, the same strain used for the BDGP's p1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at <http://bacpac.med.buffalo.edu/drosophila/bac.htm>.

FEATURES

```
1. 1920
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="RPCI-98"
/clone="BACR27A24"
```

BASE COUNT	262 a	70 c	84 g	321 t	191 others
ORIGIN	/note--end : 17-				

Query Match 5.68; Score 88.6; DB 122; Length 928;
Best Local Similarity 34.08; Pred. No. 9.6e-06;

Matches	127;	Conservative	92;	Mismatches	154;	Indels	0;	Gaps	0;
---------	------	--------------	-----	------------	------	--------	----	------	----

Qy 93 aaaatgcgagtractactaataaaaatttttacatttaattagaatttttttatcaataata 152

DB 490 AAWATAAATAATGTTAWTAAATTTTTTATKAWMAAAAAAATWTTTTTTTTTTTTTAAWAA 349

[illegible]

Db .. 610 WTAAATAWWAAAATATWTWATAATAATATWTWTWTATATATATAWAJAJAAAAA 669

Qy 273 aaCTgaaattattaataataataaactgaaaacaatttggtatcaattcatatcacatg 332

Db 670 AANWAATAAAAAATTTTAWAAWNTAAAAWAAAAAATTTTAWTTTTTTTTTTTTTTTTTTTWTATWTAA 729

Qy 333 cttagtaataaaatcgataaataattgataaactgcaaaagattttacaaaatatcttt 392

Db	730	WATAAAATATTATTTTATTTATTTGNNNNNAAAAATTTTTTAAAAATTTTTTTTTT
Qy	393	cagaaaaataataacaaattttgtcgtttctggttggtctgcaggaggatttggc

Db	790	AAAAAAAAAAAAAAAAA	WDDDDDD	WKA	KKKKKKKKKKKKKK	849
Qy	453	actatagaactct	465			

Db 850 TKTTKGARWTT 862

RESULTS	LOCUS	CNS014PQ	987 bp	DNA	GSS	26-JUL-1999
	CNS014PQ/c					
	DEFINITION	Protophila melanogaster genome survey sequence SP6 end of BAC				

ACCESSION
AL104456
VERSION
AL104456.1 GI:5616067

KEYWORDS GSS.
SOURCE fruit fly.
ORGANISM *Drosophila melanogaster*
Eubacteria. Metazoa. Arthropoda. Tracheata. Insecta.
Hymenoptera. Diptera.

REFERENCE
1 (bases 1 to 987)
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

Genoscope: Centre National de Séquençage
 TITLE Direct Submission
 JOURNAL Submitted (23-JUL-1999) Genoscope - Centre National de Séquençage

BP 191 91006 EVRY cedex - FRANCE (E-mail : segr@genoscope.cns.fr)
- web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see <http://www.fruitfly.org> The BDGP Drosophila
melanogaster BAC library was prepared by Kazutoyo Osoegawa and
Aaron Mammoser in Pieter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPCI-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain y2; cn bw sp, the same strain used for the BDGP's
P1 and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
"filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

BASE COUNT	257 a	122 c	122 g	241 t	245 others
ORIGIN					

```

/clone="BACR08K08"
/note="end : TET3"
395 a      120 c    103 g      334 t   149 others
BASE COUNT
ORIGIN
```

Query Match	5.4%	Score 85.2	DB 122	Length 1101
Best Local Similarity	40.5%	Pred. No. 3.4e-05		
Matches 223	Conservative 56	Mismatches 265	Indels 6	Gaps 1

Qy	591	aacattgagccttccaaatcatctttaaaccctcctaaacagtgtaaatgtgagacaaaataa	640
		: : : :	
Db	483	MAAAAAAAAAATTTATAAAAAAAAAAACCTTTATAAATAAAAAAAAAAAWAAAAAAAAAAAA	542
Qy	641	atattttcgaattgatagactattttttttgtgaattgacgaaaccaaacacgatt	700
		: : : :	
Db	543	AWTTTTTTWAAWTTTTTATAMATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAW	602
Qy	701	tatctggaattttagggaaccagatgtaactaaaccaatatttatttattttcttaaac	760
		: : : :	
Db	603	MATTTTTTTTTT-----TTTTTTTTTTTTTAAWTTTTTTTAAWTTTTTTAA	656
Qy	761	aaatttcatggcagcatgcctcgcccatgaaaaaaacccttaaaaaatctcacacat	820
		: : : :	
Db	657	ATAAWTTTTATTATWNNAAWATTAATAAAAAAAAAAAAAAAAAATAATAAAAAAAAAATTA	716
Qy	821	tgaccattgaaagtctgtctcccatgggtgaaccagatcaaacctcacatccaacataa	880
		: : : :	
Db	717	TAAAWTTTAAATATATATAWTTTTTAAWAAATATATTAAAAAAWAAWATWTTTTTATATTA	776
Qy	881	catgcatatctccttaccgaatcatcactaatatttttggggttaaaatatttaattcattattt	940

[illegible]

Q7 1001 1002 1003 1004 1005 1006 1007 1008 1009 1010 1011 1012 1013 1014 1015 1016 1017 1018 1019 1020 1021 1022 1023 1024 1025 1026 1027 1028 1029 1030 1031 1032 1033 1034 1035 1036 1037 1038 1039 1040 1041 1042 1043 1044 1045 1046 1047 1048 1049 1050 1051 1052 1053 1054 1055 1056 1057 1058 1059 1060 1061 1062 1063 1064 1065 1066 1067 1068 1069 1070 1071 1072 1073 1074 1075 1076 1077 1078 1079 1080 1081 1082 1083 1084 1085 1086 1087 1088 1089 1090 1091 1092 1093 1094 1095 1096 1097 1098 1099 1100 1101 1102 1103 1104 1105 1106 1107 1108 1109 1110 1111 1112 1113 1114 1115 1116 1117 1118 1119 1120 1121 1122 1123 1124 1125 1126 1127 1128 1129 1130 1131 1132 1133 1134 1135 1136 1137 1138 1139 1140 1141 1142 1143 1144 1145 1146 1147 1148 1149 1150 1151 1152 1153 1154 1155 1156 1157 1158 1159 1160 1161 1162 1163 1164 1165 1166 1167 1168 1169 1170 1171 1172 1173 1174 1175 1176 1177 1178 1179 1180 1181 1182 1183 1184 1185 1186 1187 1188 1189 1190 1191 1192 1193 1194 1195 1196 1197 1198 1199 1200 1201 1202 1203 1204 1205 1206 1207 1208 1209 1210 1211 1212 1213 1214 1215 1216 1217 1218 1219 1220 1221 1222 1223 1224 1225 1226 1227 1228 1229 1230 1231 1232 1233 1234 1235 1236 1237 1238 1239 1240 1241 1242 1243 1244 1245 1246 1247 1248 1249 1250 1251 1252 1253 1254 1255 1256 1257 1258 1259 1260 1261 1262 1263 1264 1265 1266 1267 1268 1269 1270 1271 1272 1273 1274 1275 1276 1277 1278 1279 1280 1281 1282 1283 1284 1285 1286 1287 1288 1289 1290 1291 1292 1293 1294 1295 1296 1297 1298 1299 1300 1301 1302 1303 1304 1305 1306 1307 1308 1309 1310 1311 1312 1313 1314 1315 1316 1317 1318 1319 1320 1321 1322 1323 1324 1325 1326 1327 1328 1329 1330 1331 1332 1333 1334 1335 1336 1337 1338 1339 1340 1341 1342 1343 1344 1345 1346 1347 1348 1349 1350 1351 1352 1353 1354 1355 1356 1357 1358 1359 1360 1361 1362 1363 1364 1365 1366 1367 1368 1369 1370 1371 1372 1373 1374 1375 1376 1377 1378 1379 1380 1381 1382 1383 1384 1385 1386 1387 1388 1389 1390 1391 1392 1393 1394 1395 1396 1397 1398 1399 1400 1401 1402 1403 1404 1405 1406 1407 1408 1409 1410 1411 1412 1413 1414 1415 1416 1417 1418 1419 1420 1421 1422 1423 1424 1425 1426 1427 1428 1429 1430 1431 1432 1433 1434 1435 1436 1437 1438 1439 1440 1441 1442 1443 1444 1445 1446 1447 1448 1449 1450 1451 1452 1453 1454 1455 1456 1457 1458 1459 1460 1461 1462 1463 1464 1465 1466 1467 1468 1469 1470 1471 1472 1473 1474 1475 1476 1477 1478 1479 1480 1481 1482 1483 1484 1485 1486 1487 1488 1489 1490 1491 1492 1493 1494 1495 1496 1497 1498 1499 1500 1501 1502 1503 1504 1505 1506 1507 1508 1509 1510 1511 1512 1513 1514 1515 1516 1517 1518 1519 1520 1521 1522 1523 1524 1525 1526 1527 1528 1529 1530 1531 1532 1533 1534 1535 1536 1537 1538 1539 1540 1541 1542 1543 1544 1545 1546 1547 1548 1549 1550 1551 1552 1553 1554 1555 1556 1557 1558 1559 1560 1561 1562 1563 1564 1565 1566 1567 1568 1569 1570 1571 1572 1573 1574 1575 1576 1577 1578 1579 1580 1581 1582 1583 1584 1585 1586 1587 1588 1589 1590 1591 1592 1593 1594 1595 1596 1597 1598 1599 1600 1601 1602 1603 1604 1605 1606 1607 1608 1609 1610 1611 1612 1613 1614 1615 1616 1617 1618 1619 1620 1621 1622 1623 1624 1625 1626 1627 1628 1629 1630 1631 1632 1633 1634 1635 1636 1637 1638 1639 1640 1641 1642 1643 1644 1645 1646 1647 1648 1649 1650 1651 1652 1653 1654 1655 1656 1657 1658 1659 1660 1661 1662 1663 1664 1665 1666 1667 1668 1669 1670 1671 1672 1673 1674 1675 1676 1677 1678 1679 1680 1681 1682 1683 1684 1685 1686 1687 1688 1689 1690 1691 1692 1693 1694 1695 1696 1697 1698 1699 1700 1701 1702 1703 1704 1705 1706 1707 1708 1709 1710 1711 1712 1713 1714 1715 1716 1717 1718 1719 1720 1721 1722 1723 1724 1725 1726 1727 1728 1729 1730 1731 1732 1733 1734 1735 1736 1737 1738 1739 1740 1741 1742 1743 1744 1745 1746 1747 1748 1749 1750 1751 1752 1753 1754 1755 1756 1757 1758 1759 1760 1761 1762 1763 1764 1765 1766 1767 1768 1769 1770 1771 1772 1773 1774 1775 1776 1777 1778 1779 1780 1781 1782 1783 1784 1785 1786 1787 1788 1789 1790 1791 1792 1793 1794 1795 1796 1797 1798 1799 1800 1801 1802 1803 1804 1805 1806 1807 1808 1809 1810 1811 1812 1813 1814 1815 1816 1817 1818 1

Qy 1001 gllcagacacccllccagacacg lclly lclly cllcacaagcacccllccllcllaat 1120
 : | : ||| | : | : | : | : | : | :
Db 957 TAWAANAATATTTTATTATWATWATATAWAAAAAAAATAAAATWAAATNWA 1016
 Qy 1121 acattaaaaa 1130
 | : :||| :

```

Db 1017 AWWAYAAWA 1026

```

```

RESULT 7
CNS0039G      1101 bp      DNA      GSS      03-JUN-1999
LOCUS          Drosophila melanogaster genome survey sequence TET3 end of BAC #
DEFINITION    BAC08K10 of RPCI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION     AL063921
VERSION       AL063921.1 GI:4941778
KEYWORDS      GSS.
SOURCE        fruit fly.
ORGANISM      Drosophila melanogaster
              Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
              Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
              Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE     1 (bases 1 to 1101)
AUTHORS       Direct Submission
TITLE         Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
JOURNAL       BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
              - Web : www.genoscope.cns.fr)
COMMENT       Determination of this BAC-end sequence was carried out as part of a
              collaboration with the Berkeley Drosophila Genome Project (BDGP).
              The BDGP is constructing a physical map of the Drosophila
              melanogaster genome using these BACs. For further information
              please see http://www.fruitfly.org The BDGP Drosophila
              melanogaster BAC library was prepared by Kazutoyo Osoegawa and
              Aaron Mammosser in Pieter de Jong's laboratory in the Department of
              Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
              NY. The library is named RPCI-98 and was constructed by partial
              EcoRI digestion of Drosophila DNA provided by the BDGP from the
              isogenic strain y2; cn bw sp, the same strain used for the BDGP's
              p1 and EST libraries. A more detailed description of the library
              and how to order individual BAC clones, the entire library, or
              filters for hybridization from the BACPAC Resource Center can be
              found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
              Location/Qualifiers
              1. .1101
              /organism="Drosophila melanogaster"
              /db_xref="taxon:7227"
              /clone_lib="RPCI-98"
              /clone="BAC08K10"
              /note="end : TET3"

BASE COUNT    201 a 64 c 131 g 202 t 503 others
ORIGIN

Query Match      5.3%  Score 84;  DB 122;  Length 1101;
Best Local Similarity 18.5%  Pred. No. 5.4e-05;
Matches 129;  Conservative 292;  Mismatches 272;  Indels 3;  Gaps 1;

QY 249 aatcttatgtcttcaaatgagactgaaattattattattataactgaaacaa 308
      ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
DB 402 ANWWTTTTTTTAAWAAWAAWAAATTTWAAWAAWAAWAAATTTWAAWAAWAAWAAWTA 461
      ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :

QY 309 atttggtacattcatatcatcattgtagtaataatgcgataattatgataaattc 368
      ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
DB 462 WNTTTAAWAAAAAATAATTTTWTTTTAAWAAWAAWAAWTTTAAWAAWAAWAA 521
      ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :

QY 369 gcaaaagatttcacaaattcttcagaaataaatt---aacaaatttctgctttc 425
      ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
DB 522 AAAAAAATAAATAAATAATTTTAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAW 581
      ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :

QY 426 atggtgtgtggtgagagattggtgacattgacattcctcagcaccattcttgc 485
      ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
DB 582 YTYTYTYTYTYTYTYTYTYTYTYTYTYTYTYTYTYTYTYTYTYTYTYTYTYTYTY 641
      ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :

QY 486 ctctcaactaaacagtgctcagattggtggtgatttatcatcagcattcccttcaa 545
      ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
DB 642 YTYTYTYTYTYTYTYTYTYTYTYTYTYTYTYTYTYTYTYTYTYTYTYTYTYTYTY 701
      ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :

QY 546 aacttcttacttctgctgctggtggtggtggtggtggtggtggtggtggtggtggt 605
      ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
DB 702 YYYHHYTAHHTHTHHWAAHYHYWYTYWAAWYWYMYCTACTYTHHHHHHHYHWTWYAWA 761

```

```

QY 606 taacccttaaacagtaaatgtgaaggacacaaataatttttcaaatgttagagatat 665
      ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
DB 762 HAMMMHHHAHYAAAAWAAWAAWTAHTHHYHHYTHHHYHHYHYHYMYTCCYMCYHCHYHT 821
      ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :

QY 666 ttttttttgaatttgagacacacacacacacacacacacacacacacacacacacaga 725
      ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
DB 822 AYTCTWTHTHHMTWTTTHYHTHHHTTTTHAAWAAWAAWTTTTCWMMWAAWTTTATWATHCWACMT 881
      ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :

QY 726 tgaactaaacacacacacacacacacacacacacacacacacacacacacacacag 785
      ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
DB 882 MHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH 941
      ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :

QY 786 cccatgaaaaaaccttataaaatatctacacattgacacattgaaagtctgtctccc 845
      ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
DB 942 MATWTTTMTTMMCMCHHHCHYHHMMHYMYCCHYUCTCTHTATTHYHMYCTCYHCT 1001
      ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :

QY 846 atgggtaaacacacacacacacacacacacacacacacacacacacacacacacata 905
      ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
DB 1002 WHTYTYTAWTAHMTTATWMMWMMWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAW 1061
      ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :

QY 906 ctaattatttgggttaataataataataataataataataataataataataataata 941
      ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
DB 1062 CWWYHHHTTHMMWMMWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAW 1097
      ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :

RESULT 8
CNS00CS1/c      843 bp      DNA      GSS      04-JUN-1999
LOCUS          Drosophila melanogaster genome survey sequence TET3 end of BAC #
DEFINITION    BAC26H19 of RPCI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION     AL059666
VERSION       AL059666.1 GI:4947129
KEYWORDS      GSS.
SOURCE        fruit fly.
ORGANISM      Drosophila melanogaster
              Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
              Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
              Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE     1 (bases 1 to 843)
AUTHORS       Direct Submission
TITLE         Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
JOURNAL       BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
              - Web : www.genoscope.cns.fr)
COMMENT       Determination of this BAC-end sequence was carried out as part of a
              collaboration with the Berkeley Drosophila Genome Project (BDGP).
              The BDGP is constructing a physical map of the Drosophila
              melanogaster genome using these BACs. For further information
              please see http://www.fruitfly.org The BDGP Drosophila
              melanogaster BAC library was prepared by Kazutoyo Osoegawa and
              Aaron Mammosser in Pieter de Jong's laboratory in the Department of
              Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
              NY. The library is named RPCI-98 and was constructed by partial
              EcoRI digestion of Drosophila DNA provided by the BDGP from the
              isogenic strain y2; cn bw sp, the same strain used for the BDGP's
              p1 and EST libraries. A more detailed description of the library
              and how to order individual BAC clones, the entire library, or
              filters for hybridization from the BACPAC Resource Center can be
              found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
              Location/Qualifiers
              1. .843
              /organism="Drosophila melanogaster"
              /db_xref="taxon:7227"
              /clone_lib="RPCI-98"
              /clone="BAC26H19"
              /note="end : TET3"

BASE COUNT    131 a 102 c 259 g 141 t 210 others
ORIGIN

Query Match      5.3%  Score 83.8;  DB 122;  Length 843;

```

```

Best Local Similarity 39.2%; Pred. NO. 6e-05;
Matches 122; Conservative 64; Mismatches 125; Indels 0; Gaps 0;

Qy 851 taaccagatcaaacacacatacaatacaatgagatctctcttaccacataactaat 910
  || || || || || || || || || || || || || || || || || || || ||
Db 833 TATAWATAWATAWATAWATAWATAWATAWATAWATAWATAWATAWATAWATAW 774
  || || || || || || || || || || || || || || || || || || || ||
Qy 911 ttttttgggttaaaatcattatttatttatttatttatttatttatttatttatt 970
  || || || || || || || || || || || || || || || || || || || ||
Db 773 TATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTAT 714
  || || || || || || || || || || || || || || || || || || || ||
Qy 971 ttttaaaaaatgataaaatattattattattattattattattattattattatt 1030
  || || || || || || || || || || || || || || || || || || || ||
Db 713 ATATTATATATATATATATATATATATATATATATATATATATATATATATAT 654
  || || || || || || || || || || || || || || || || || || || ||
Qy 1031 aaatatatttttttatttttttatttttttatttttttatttttttattttttt 1090
  || || || || || || || || || || || || || || || || || || || ||
Db 653 AWTTTTWTATWATAWATAWATAWATAWATAWATAWATAWATAWATAWATAWATA 594
  || || || || || || || || || || || || || || || || || || || ||
Qy 1091 tctgtttcaaaaagcattcatttatttatttatttatttatttatttatttatt 1150
  || || || || || || || || || || || || || || || || || || || ||
Db 593 AAATWATATATAAAWATAWATAWATAWATAWATAWATAWATAWATAWATAWATA 534
  || || || || || || || || || || || || || || || || || || || ||
Qy 1151 aatcttcttgt 1161
  || || || || || || || || || || || || || || || || || || || ||
Db 533 WAATWATWAT 523

RESULT 9
CNS0021J/c CNS0021J 1101 bp DNA GSS 03-JUN-1999
LOCUS Drosophila melanogaster genome survey sequence TET3 end of BAC #
DEFINITION BACR03N11 of RPCI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL061936.1 GI:4940214
VERSION AL061936
KEYWORDS GSS.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1101)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see http://www.fruitfly.org The BDGP Drosophila
melanogaster BAC library was prepared by Kazutoyo Osogawa and
Aaron Mammeter in Pletier de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPCI-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain y2; cn bw sp, the same strain used for the BDGP's
P1 and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
Location/Qualifiers
1. 1101
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone-lib="RPCI-98"
/clone="BACR03N11"
/note="end : TET3"

BASE COUNT 631 a 7 c 28 g 289 t 146 others
ORIGIN

```

```

Query Match 5.3%; Score 83; DB 122; Length 1101;
Best Local Similarity 40.2%; Pred No. 7.9e-05;
Matches 210; Conservative 61; Mismatches 245; Indels 6; Gaps 1;

Qy 740 tatttatttttttctaaacaaaatttctgagcagcatgcctcagcccatgataaaac 799
  || || || || || || || || || || || || || || || || || || || ||
Db 1043 TTTTTTTTTTTTTTTTNNMAAATATTTCACATTTTTCATNCYTCWTATTTTTCMTWH. 984
  || || || || || || || || || || || || || || || || || || || ||
Qy 800 cttataaaaatcctacacattgacattgaaaagtctctccatcgggtaaccagat 859
  || || || || || || || || || || || || || || || || || || || ||
Db 983 TTTTITAMMATTMAAMMATTATYMTCTTACHATTYTTAAACMYCMMYTAMCMHMCATTC 924
  || || || || || || || || || || || || || || || || || || || ||
Qy 860 caactccatcccaacataacatgatatctctccaacatcataactattcatt---- 914
  || || || || || || || || || || || || || || || || || || || ||
Db 923 MCATWTTTAAWACTTTAAACAAACATATTTAAATTSATTTTAAATTTTAAATAAATA 864
  || || || || || || || || || || || || || || || || || || || ||
Qy 915 -ttgggttaaaatcattcatttatttatttatttatttatttatttatttatttatt 973
  || || || || || || || || || || || || || || || || || || || ||
Db 863 ATTTAAAAAAAWTTTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWT 804
  || || || || || || || || || || || || || || || || || || || ||
Qy 974 aaaaaaatgataaaaattattattattattattattattattattattattattatt 1033
  || || || || || || || || || || || || || || || || || || || ||
Db 803 ATTAATAAAATTAATAATTAATAATTAATAATTAATAATTAATAATTAATAATTA 744
  || || || || || || || || || || || || || || || || || || || ||
Qy 1034 tatatttttttttatttatttatttatttatttatttatttatttatttatttatt 1093
  || || || || || || || || || || || || || || || || || || || ||
Db 743 TWTTTTAAWATAWATAWATAWATAWATAWATAWATAWATAWATAWATAWATA 684
  || || || || || || || || || || || || || || || || || || || ||
Qy 1094 tgtttacaagaagcattcattcattcattcattcattcattcattcattcattcatt 1153
  || || || || || || || || || || || || || || || || || || || ||
Db 683 WATWATAAAWAAWTTTWTATTTTWTATTTTAAWAAAAAAWAAWTTTAAWAAW 624
  || || || || || || || || || || || || || || || || || || || ||
Qy 1154 ctctctggagtggtggtggtggtggtggtggtggtggtggtggtggtggtggt 1213
  || || || || || || || || || || || || || || || || || || || ||
Db 623 TTTTTTTTWTATWATAWATAWATAWATAWATAWATAWATAWATAWATAWATAWATA 564
  || || || || || || || || || || || || || || || || || || || ||
Qy 1214 aaccagaagacaaaataaagtagtcacacacacacacacacacacacacacac 1255
  || || || || || || || || || || || || || || || || || || || ||
Db 563 TTAWAAAAAAWAAAAWAAAAWAAAAWAAAAWAAAAWAAAAWAAAAWAAAAWATA 522
  || || || || || || || || || || || || || || || || || || || ||

RESULT 10
CNS016LI/c CNS016LI 1101 bp DNA GSS 26-JUL-1999
LOCUS Drosophila melanogaster genome survey sequence T7 end of BAC
DEFINITION BACN16D22 of DrosBAC library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL106896
VERSION AL106896.1 GI:5624374
KEYWORDS GSS.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1101)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a
collaboration with the European Drosophila Genome Project (EDGP) -
http://www.edgp.ebi.ac.uk - This Drosophila melanogaster BAC
library (Dros BAC) was made by Alain Billaud at CSH (Centre
d'Etude du Polymorphisme Humain) with funding provided by a MRC
project grant. The DNA was prepared from embryos by Alain Bucheton
and Genevieve Payan. It has been constructed in the vector
pBelOBAC11.
Location/Qualifiers
1. 1101
/organism="Drosophila melanogaster"

```


project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBelobAC11.

FEATURES

```
/organism="Drosophila melanogaster"  
/plasmid="pBelobAC11"  
/db_xref="taxon:7227"  
/clone_lib="DrosBAC"  
/clone="BACN15C13"  
/note="end : sp6"
```

BASE COUNT	448 a	136 c	161 g	336 t	120 others
ORIGIN					

Query Match 5.2%; Score 81.2; DB 123; Length 1201;
Best Local Similarity 44.8%; Pred. No. 0.00016;
Matches 137; Conservative 46; Mismatches 121; Indels 2; Gaps 1;

Oy 860 caaactcacatccaaacataaaccatggaatatctccttaaccaatcatactaatattttgga 919

Oy 860 caaactcacatccaacaacatacgaatatctcctttaccaatcatactaatatttttggg 919

[illegible]

DB 1190 M M M M A M M M M M M H M T C T T T T T T T A T A T A A A A A A A W A T T W T T T T T W T T T T T T T T A A W C 1131

Qy 920 ttaaataatcattatttttaagatatataatgaagattttttaaaaa 979

1130 TTYTWTATTAATWTAWAATTTWATTTATAWAWAAWATTAATAHAAHTTTTTTTTTAMTAWA 1071

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

Qy 980 atgtataaaattatatattcatgatttttcatcacatttgatttgataataa--atata 1037

Db 1070 ATTTTWAATTTTATTTTAAATTTTATTTTAAATAWTTTTT 1011

[illegible]

Af: .

Db 1010 TTTTATTAATTATTAATAAAATWMAATAAATAATT 951

Qy 1098 taaaaagcattcatcttaataacattaaaaaatatttaatactaacagtagaatcttc 1157

[illegible]

530 TTTWAGAGAAWTAAGAAWWTTTTITTAWAATMCMAATWWINIAAAAAATAATAATWWTATTT 891

QY 1158 ttgtga 1163

890 TTNTAA 885

RESULT 14

CNS0145P

LOCUS	CNS0145P	1043 bp	DNA	GSS	26-JUL-1999
DEFINITION	Proscophila melanogaster genome survey sequence T7 end of PAC				

BACN11G11 of DrosBAC library from *Drosophila melanogaster* (fruit

ACCESSION AY102725
fly), genomic survey sequence.

AL103735.1 GI:5615346

KEYWORDS
GSS,
50044, 50045

SOURCE	FRUIT FLY.
ORGANISM	<i>Protophila melanogaster</i>

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila

REFERENCE
1 (bases 1 to 1043)

AUTHORS	TITLE	Genoscope. Direct Submission
...

JOURNAL
Submitted (23-JUL-1999) Genoscope - Centre National de Sequençage

BP 191 91006 EVRY cedex - FRANCE (E-mail : secref@genoscope.cnrs.fr)

web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part

collaboration with the European Drosophila Genome Project (EDGP)

<http://www.edgip.ebi.ac.uk>. This *Drosophila melanogaster* BAC library (Dros BAC) was made by Alain Billaud at CEPH (Centre

d'Etude du Polymorphisme Humain) with funding provided by a MRC

project grant. The DNA was prepared from embryos by Alain Buche and Genevieve Pavan. It has been constructed in the vector

pBelovAC11.

[illegible]


```

source
1..1043
/organism="Drosophila melanogaster"
/plasmid="pBelobAC11"
/db_xref="taxon:7227"
/clone_lib="DrosBAC"
/clone="BACN15E04"
/note="end : T7"

BASE COUNT 277 a 96 c 121 g 382 t 167 others
ORIGIN

Query Match 5.1%; Score 81; DB 123; Length 1043;
Best Local Similarity 41.2%; Pred. No. 0.00017;
Matches 153; Conservative 58; Mismatches 158; Indels 2; Gaps 1;

QY 887 tatctccctaccacataactatttttgggttcaaatcaatcatttttcaagat 946
DB 548 TCCTTTTWTATATATATATATATATATATATATATATATATATATATATWA 607
QY 947 attaatgaagaattaaagattttttaaataataataataataataataatgat 1006
DB 608 AAAAAATATATATATATATATATATATATATATATATATATATATATTTWT 667
QY 1007 ttctacataattgattgataataataataatttttttaatttttaaaaaatgttgc 1066
DB 668 TTTTWTATATATATATATATATATATATATATATATATATATATATATTTWA 727
QY 1067 agacacttattgacatgctgttctgtttcaaaaagcattcatttcaatcatta 1126
DB 728 ATTTATATATTTTAAATATATATATATATATATATATATATATATATATATWA 787
QY 1127 aaaaatattt--aataactaacagtagaactcttctgtgagtggtgaggagagcaacc 1184
DB 788 AAAAAWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 847
QY 1185 tggcattgaacgagaagagagtagcagacacagaagacaaataaaaagtagtgcacaa 1244
DB 848 TAAWAWAAAAAATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAHAYATACAAAA 907
QY 1245 acaaatcaaaa 1255
DB 908 MCWATHAWAA 918

RESULT 15
CNS016CO/1200 bp DNA GSS 26-JUL-1999
Drosophila melanogaster genome survey sequence T7 end of BAC
BACN15E04 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL106578
VERSION AL106578.1 GI:5622626
KEYWORDS GSS.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1200)
Genoscope.
Direct Submission
Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
collaboration with the European Drosophila Genome Project (EDGP) -
http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
library (Dros BAC) was made by Alain Billaud at CEPH (Centre
d'Etude du Polymorphisme Humain) with funding provided by a MRC
project grant. The DNA was prepared from embryos by Alain Bucheton
and Genevieve Payan. It has been constructed in the vector
pBelobAC11.
Location/Qualifiers

```

```

source
1..1200
/organism="Drosophila melanogaster"
/plasmid="pBelobAC11"
/db_xref="taxon:7227"
/clone_lib="DrosBAC"
/clone="BACN15E04"
/note="end : T7"

BASE COUNT 351 a 188 c 167 g 307 t 187 others
ORIGIN

Query Match 5.1%; Score 80; DB 123; Length 1200;
Best Local Similarity 38.8%; Pred. No. 0.00025;
Matches 132; Conservative 61; Mismatches 147; Indels 0; Gaps 0;

QY 94 aaatgcgagtaactaaataaaattttacatttaaatagaaatttttttcaataaataat 153
DB 1199 ATATTWATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 1140
QY 154 taatttattagttttattagaataatttaattagaataattttgaatcccgatttctc 213
DB 1139 WAATWTATTTATTTTAAATATTTTAAATATTTTAAATATTTTAAATATTTTAAAT 1080
QY 214 cttttcttcgctattcattcattttctaaacaaacaaatcttatgttcttcaaataga 273
DB 1079 TTTTWTATTTTAAATATTTTAAATATTTTAAATATTTTAAATATTTTAAATATTT 1020
QY 274 acttgaataatttaattataataactgaacaaacttttgggtatcaatcattacatgc 333
DB 1019 AWTTTTATTTTAAATATTTTAAATATTTTAAATATTTTAAATATTTTAAATATTT 960
QY 334 ttgataataaatgcgataaatttaattgataaattctgcaaaagatttttcaataatcttc 393
DB 959 TAAATTTAAATAAAATTTTAAATATTTTAAATATTTTAAATATTTTAAATATTT 900
QY 394 agaaaaaatttaatacaaaatttttgcgttttcatggggtt 433
DB 899 TCHTTATTTTAAATAAAATTTTAAATATTTTAAATATTTTAAATATTTTAAATATTT 860

Search completed: September 2, 2000, 00:25:15
Job time: 4339 sec

```

Result No.	Score	Query Match	Length	DB	ID	Description
1	57.2	3.6	3095	7	5231168-1	Patent No. 5231168-1
2	56	3.6	19124	4	US-08-487-886B-13	Sequence 13, Appl
3	54.2	3.4	1186	4	US-08-731-722-5	Sequence 5, Appl
4	53.8	3.4	6243	1	US-09-056-075-1	Sequence 1, Appl
5	53.2	3.4	3138	1	US-07-867-106-4	Sequence 4, Appl
6	52.6	3.3	665	4	US-08-883-795A-36	Sequence 36, Appl
7	52	3.3	8920	3	US-08-446-885A-1	Sequence 1, Appl
8	51.4	3.3	19124	4	US-08-487-886B-13	Sequence 13, Appl
9	51.2	3.3	665	4	US-08-883-795A-36	Sequence 36, Appl
10	51.2	3.3	3933	4	US-08-731-722-3	Sequence 3, Appl
11	51.2	3.3	3933	4	US-08-731-722-3	Sequence 3, Appl
12	50.6	3.2	8920	3	US-08-446-885A-1	Sequence 1, Appl
13	50.2	3.2	473	2	US-08-764-100-16	Sequence 16, Appl
14	50.2	3.2	4970	2	US-08-764-100-14	Sequence 14, Appl
15	50.2	3.2	4970	2	US-08-764-100-20	Sequence 20, Appl
16	49.8	3.2	4467	2	US-08-565-907A-1	Sequence 1, Appl
17	49.8	3.2	4467	4	US-08-910-551B-1	Sequence 1, Appl
18	49.8	3.2	4467	4	US-08-909-425A-1	Sequence 1, Appl
19	49.4	3.1	1611	7	5213972-6	Patent No. 5213972
20	49.2	3.1	319	1	US-07-593-657-14	Sequence 14, Appl
21	49	3.1	4098	4	US-08-605-106-4	Sequence 4, Appl
22	49	3.1	6768	2	US-08-107-755A-1	Sequence 1, Appl
23	49	3.1	8457	1	US-07-991-867B-1	Sequence 1, Appl
24	49	3.1	8457	4	US-08-544-332-1	Sequence 1, Appl
25	48.4	3.1	1415	1	US-08-413-118-126	Sequence 126, App
26	48.4	3.1	1415	5	US-08-473-446-126	Sequence 126, App

QY 1083 ta 1084
Db 2953 TA 2954

RESULT 6
US-08-883-795A-36
; Sequence 36, Application US/08883795A
; Patent No. 5985607
; GENERAL INFORMATION:
; APPLICANT: Delcuve, Genevieve
; APPLICANT: Awang, Gregor
; TITLE OF INVENTION: Recombinant DNA Molecules and Expression
; TITLE OF INVENTION: Vectors for Tissue Plasminogen Activator
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BERESKIN & PARR
; STREET: 40 King Street West
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5H 3Y2
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/883,795A
; FILING DATE: 27-JUN-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Gravelle, Micheline
; REGISTRATION NUMBER: 40,261
; REFERENCE/DOCKET NUMBER: 7841-062
; TELEPHONE: (416) 364-7311
; TELEFAX: (416) 361-1398
; INFORMATION FOR SEQ ID NO: 36:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 665 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; ORGANISM: Homo sapiens
; IMMEDIATE SOURCE:
; CLONE: Rh 32
; US-08-883-795A-36

Query Match 3.3%; Score 52.6; DB 4; Length 665;
Best Local Similarity 55.1%; Pred. No. 0.033;
Matches 124; Conservative 0; Mismatches 99; Indels 2; Gaps 1;
QY 925 tattaatcattatttttaagatatttaagaataatttaagaatttttttaaaaaaatgta 984
Db 25 TATAATTAATATTTTATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTT 84
QY 985 taaattattatttcattgatttttcacatttgatttgataataataatattttttt 1044
Db 85 TATAATTAATTAATTTTATAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 142
QY 1045 taattttcttaaaaaattgcaagacacttattagacatgctctgtctgtttacaaaa 1104
Db 143 TTTATAATTAATTAATTTTATAATTAATTAATTTTATAATTAATTAATTAATTAATAT 202
QY 1105 gcattcatcatttaatacaataaaaaatttttaatacaaacagta 1149
Db 203 TTTATAATTAATTAATTTTATAATTAATTAATTTTATAATTAATTAATTAATTAATA 247

RESULT 7
US-08-446-855A-1/c
; Sequence 1, Application US/08446855A
; Patent No. 5849573
; GENERAL INFORMATION:
; APPLICANT: Stewart, Thomas S
; APPLICANT: Flores, Maria V
; APPLICANT: O'Sullivan, William J
; TITLE OF INVENTION: Nucleotide sequence encoding carbamoyl
; TITLE OF INVENTION: phosphate synthetase II
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon & Vanderhye PC
; STREET: 1100 NO. 5849573th Glebe Road, 8th Floor
; CITY: Arlington
; STATE: Virginia
; COUNTRY: USA
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.24
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/446,855A
; FILING DATE: 06-Jul-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Mitchard, Leonard C
; REGISTRATION NUMBER: 29,009
; REFERENCE/DOCKET NUMBER: 47-80
; TELEPHONE: 703-816-4000
; TELEFAX: 703-816-4100
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8920 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: genomic
; US-08-446-855A-1

Query Match 3.3%; Score 52; DB 3; Length 8920;
Best Local Similarity 49.6%; Pred. No. 0.06;
Matches 133; Conservative 0; Mismatches 135; Indels 0; Gaps 0;
QY 903 atactaattttttgggttaataatttaattcatttttttaagatatatttaagaataa 962
Db 777 ATATTTATATATATATCTATATTTTCCCATTTTCTTTTATATACATTT 718
QY 963 aagatttttttaaaaaatgtataaaattattattcattgatttttcacacatttgatt 1022
Db 717 ATATTTATGTTTAAATAATTTATTAATTTTACATATACATACAGTTTCATTTTCA 658
QY 1023 ttgataataaatatatttttttcaattttcttaaaaagtgtgcaagacatttagaca 1082
Db 657 TTTTCTTTTCTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 598
QY 1083 tagcttctgtctgttttacaagaagcattcatttaatacaataaaaaattttaact 1142
Db 597 TTTATAAACTAAGAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 538
QY 1143 aacagttagaatctcttctgtgagtggtgt 1170
Db 537 AATTATAGGAACCAACATATATGGGGACT 510

RESULT 8
US-08-487-826B-13/c
; Sequence 13, Application US/08487826B
; Patent No. 5993827

;; TITLE OF INVENTION: Nucleotide sequence encoding carbamoyl
;; TITLE OF INVENTION: phosphate synthetase II
;; NUMBER OF SEQUENCES: 2
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Nixon & Vanderhye PC
;; STREET: 1100 No. 5849573th Glebe Road, 8th Floor
;; CITY: Arlington
;; STATE: Virginia
;; COUNTRY: USA
;; ZIP: 22201-4714
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Patent In Release #1.24
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/446,855A
;; FILING DATE: 06-Jul-1995
;; CLASSIFICATION: 435
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Mitchard, Leonard C
;; REGISTRATION NUMBER: 29,009
;; REFERENCE/DOCKET NUMBER: 47-80
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 703-816-4000
;; TELEFAX: 703-816-4100
;; INFORMATION FOR SEQ ID NO: 1:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 8920 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: genomic
;; US-08-446-855A-1

Query Match 3.2%; Score 50.6; DB 3; Length 8920;
Best Local Similarity 47.6%; Pred. No. 0.11;
Matches 149; Conservative 0; Mismatches 164; Indels 0; Gaps 0;

QY 792 aaaaaaaccttataaaaaatctacacattgacattgaaaggcttcctccatgggt 851
DB 8513 ATAATACGTTTAAA 8572

QY 852 aaccagatcaaacatccacataacatgatatccctaccatacataaatt 911
DB 8573 TATTATAAATGTTTTTATTAAATATGATCATTAATTTATATGATTTATTTTATAA 8632

QY 912 attttgggttaaatattacattatttttaagatatatttaagaataataaagatttt 971
DB 8633 ATTTTGTGTATATATACAAATTTTATTATTCACATCATATGATATAACCAAAATGGTTT 8692

QY 972 ttaaaaaatgtaaaaaattattatttcattgatttttcacattgatttgataata 1031
DB 8693 TTCAATTTTACAAATTTTATAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 8752

QY 1032 aatatatttttttaatttttttaaaaaattgtcgaagacacttattagacatgcttgt 1091
DB 8753 TATATAAACATTAATAATGATATAAATCTTTTAAATTTATATATATATATATAAATGTTATGA 8812

QY 1092 tctgtttacaaaa 1104
DB 8813 TTTTAAAAA 8825

RESULT 13
US-08-764-100-16
; Sequence 16, Application US/08764100
; Patent No. 5773700
; GENERAL INFORMATION:
; APPLICANT: van Grinsven J., Martinus Q.
; APPLICANT: De Haan, Petrus T.
; APPLICANT: Gielen L., Johannes J.

;; APPLICANT: Peters, Dirk
;; APPLICANT: Goldbach, Robert W.
;; TITLE OF INVENTION: Improvements in or Relating to Organic
;; TITLE OF INVENTION: Compounds
;; NUMBER OF SEQUENCES: 27
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Sandoz Agro, Inc.
;; STREET: 975 California Avenue
;; CITY: Palo Alto
;; STATE: CA
;; COUNTRY: USA
;; ZIP: 94304
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Patent In Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/764,100
;; FILING DATE: 06-DEC-1996
;; CLASSIFICATION: 800
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US/08/214,064
;; FILING DATE:
;; APPLICATION NUMBER: US 08/032,235
;; FILING DATE: 17-MAR-1993
;; APPLICATION NUMBER: GB 9206016.9
;; FILING DATE: 19-MAR-1992
;; ATTORNEY/AGENT INFORMATION:
;; NAME: No. 5773700rils, Allen E.
;; REGISTRATION NUMBER: 34,490
;; REFERENCE/DOCKET NUMBER: 137-1061
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (415) 354-3592
;; TELEFAX: (415) 857-1125
;; INFORMATION FOR SEQ ID NO: 16:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 473 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; US-08-764-100-16

Query Match 3.2%; Score 50.2; DB 2; Length 473;
Best Local Similarity 52.1%; Pred. No. 0.09;
Matches 112; Conservative 0; Mismatches 103; Indels 0; Gaps 0;

QY 922 aaattattacattatttttaagatatatttaagaataataaagatttttttaaaaaat 981
DB 236 AARACCAAAATAATTTTGTGTAATAATAAGCTCCGCCAGATTGTGCTAAGACC 295

QY 982 gtataaaattatattatcattgatttttcacattgatttttgataataaataatttt 1041
DB 296 TTTTATTGTTTATACATTTTATTGTTTGTGCAATTTTATTATTATTATTATTATTTA 355

QY 1042 ttttaatttttttaaaaaattgtcgaagacacttattagacatgcttctgttttaca 1101
DB 356 TATTTTATATAGTTGCTTATTATTAACATTATTATTAACAAATTAATTTATTATTGATTA 415

QY 1102 aagcattcatcatttaataacattaaaaaatttt 1136
DB 416 CAATCATTCGCTTATTATTAATTTAAACACATTT 450

RESULT 14
US-08-764-100-14
; Sequence 14, Application US/08764100
; Patent No. 5773700
; GENERAL INFORMATION:
; APPLICANT: van Grinsven J., Martinus Q.
; APPLICANT: De Haan, Petrus T.
; APPLICANT: Gielen L., Johannes J.

APPLICANT: Peters, Dirk
 APPLICANT: Goldbach, Robert W.
 TITLE OF INVENTION: Improvements in or Relating to Organic
 TITLE OF INVENTION: Compounds
 NUMBER OF SEQUENCES: 27
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Sandoz Agro, Inc
 STREET: 975 California Avenue
 CITY: Palo Alto
 STATE: CA
 COUNTRY: USA
 ZIP: 94304
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/764,100
 FILING DATE: 06-DEC-1996
 CLASSIFICATION: 800
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US/08/214,064
 FILING DATE:
 APPLICATION NUMBER: US 08/032,235
 FILING DATE: 17-MAR-1993
 APPLICATION NUMBER: GB 9206016.9
 FILING DATE: 19-MAR-1992
 ATTORNEY/AGENT INFORMATION:
 NAME: No. 5773700119, Allen E.
 REGISTRATION NUMBER: 34,490
 REFERENCE/DOCKET NUMBER: 137-1061
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (415) 354-3592
 TELEFAX: (415) 857-1125
 INFORMATION FOR SEQ ID NO: 20:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 4970 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 US-08-764-100-20

Query Match 3.28; Score 50.2; DB 2; Length 4970;
Best Local Similarity 52.18; Pred. No. 0.12;
Matches 112: Conservative 0; Mismatches 103: Indels 0; Gaps 0;

922	aaataattaatcattattttttaagatataataattaagaataattaaagaatgtttttttaaaaaaat	981
Qy		
3738	aaacacaaaaaaatTTTTTTTGTAAATAAATAAGCGTCGCGCCAGATTTTGGTCTAAGACC	3679
Ddb		
982	gtataaaaattattattcatgatttttcatacatattgatttgataataaatatatttt	1041
Qy		
3678	TTTTTATTGTGTTTTTATACATTTATTGTGTTGTTGATTTTTATTATTATTATTTTA	3619
Ddb		
1042	ttttaatttcttaaaaaatgftcgaagacacattatagacatagcttgctcgtgttaca	1101
Qy		
3618	TATTTTTTATAGTTTGCATTATTATACACTTATTAGACAAATTAATTTATTTCATTA	3559
Ddb		
1102	aaagcattcatcatttaatacatcaaaaaatattt	1136
Qy		
3558	CAATCATCTCGCTATTTTTAATTTAAACACACATTT	3524
Ddb		

Search completed: September 2, 2000, 02:00:43
Job time: 8407 sec

[illegible]

RESULT	4
N71405/G	
ID	N71405 standard; DNA; 1864 BP.
AC	N71405;
DT	18-APR-1991 (first entry)
DE	Sequence of ANS-1 which increases transformation efficiency.
EN	Enzyme; fungal expression vector; Aspergillus expression vector;
KW	Trichoderma; ds.

OS Mucor miehei.
 PN EP-215594-A.
 PD 25-MAR-1987.
 PP 27-AUG-1986; 306624.
 PR 29-AUG-1985; US-771374.
 PR 07-JUL-1986; US-882224.
 PR (GENE-) GENENCOR INC.
 PI Cullen D, Gray GL, Hayenga KJ, Lawlis VB;
 PD WPI; 87-095049/14.
 PR New DNA sequences for expressing polypeptide in filamentous fungi
 PT - with secretion of prod. from the cells, and new vectors and
 PT transformed fungi.
 PT Example; Fig 13; 45pp; English.
 PS A DNA sequence coding for a heterologous polypeptide which can be
 CC expressed in and secreted from filamentous fungi is claimed. Pref.
 CC the DNA sequence codes for bovine preprochymosin, M. miehei
 CC preprocarboxyl protease or A. niger preproglucanase. Also new
 CC are vectors consisting of the DNA sequence plus an operably-linked
 CC signal sequence. The vectors may also include a sequence which
 CC increases transformation efficiency, e.g. ANS-1.
 CC Sequence 1864 BP; 786 A; 210 C; 44 G;
 SQ Sequence 1864 BP; 786 A; 210 C; 44 G; 732 T;

	Query Match.	3.78;	Score 59;	DB 1;	Length 1864;
	Best Local Similarity	42.89;	Pred. No. 0.09;		
	Matches 452;	Conservative 0;	Mismatches 583;	Indels 22;	Gaps 5;
Qy	126	aattcagaattttttatccaataataattattattagttttatttagaataattataa	185		
Db	1554	ATTTCCTAATNTNTTAGGAGAGNTATCTATCTCTTAAATCATCTTAATAATTTTTTATA	1495		
Qy	186	gaaaattttgaaatccccgatttctctctcttttcttcgctattctatcttaaccaa	245		
Db	1494	GAGNTTTATTATTNNNNTAGAGTAGTATTTTTTATAGTTCTCTTAAGTAAATTAATTTTAGNTT	1435		
Qy	246	accaattctatatgtcttccaataatagaaactgaaattattataattataataaactcgaa-	304		
Db	1434	AGTTATTATAATTAATAATATTATAATTAAGTTAGTAANAGTTTNAAGTGCAAAATATAT	1375		
Qy	305	-aacaatttggtatcaattctatacacatgccttagtaataaaatgcgataaattatgtata	363		
Db	1374	TTATATATTTTAAATTAATTAATTTAAATNTTTAGTATTANTAAGATTTAAGTTATAGTTA	1315		
Qy	364	aatctgcaaaagattttacaataatctcttcagaaaaaattataaacaattttgcggtt	423		
Db	1314	GATAAAATCTATTNTTAAATATATTATTNAGTTTGTAGTAGTATTTTTTTATATATTTTAAAGT	1255		
Qy	424	tcatgggtgtggctcgaggaggatttgggcactatagaactctcctcaacggaccattcttgg	483		
Db	1254	TAAATATTGTNTANNNTGGNTAGTATATCTTATTATATAGTATAATAATATATNTAGGNA	1195		
Qy	484	cacttcaactaaacgattgctcagaattggtggggattttatat--tcaagcatatccctt	541		
Db	1194	AGNTGCTAGCTTANGTAGTTAAATATTTTAAAAATAAAATTTAAATAATTAAGTATATCTGTN	1135		

Qy	663	tatttttttggtaattgagcaaaaaccaggattatcctcgaaatttttagaacccac	722
Db	1664		
		TTTTTTTTTTTTTTTTTTT-----TTAACCGAATGAATAATATTATGAGATCAC	1715
Qy	723	agatgaactcaacaataattattatttttctaatacaaaatttcctgcagcatgcct	782
Db	1716		
		AATTTTAAATAAT--TTATTTTTTTTCGGGTTAATATTTTTTTTGTCACGATAACT	1772
Qy	783	cagcccatgaaaaaaccttataaaaatatctcacacattgaccttgaaaagtctgcgtct	842
Db	1773		
		AATCCCTTG---TTATCTTAAAAAAAATGTACACATGTACAATATGTATATATATA	1829
Qy	843	cccatgggtaaccagatcaaatccacatcccaacaataacatcgatctccttaccaatc	902
Db	1830		
		TATATATATGTATCTTACAATTTAATAATATTTTAGAATATTAATAATTAATCTTAGAA	1889
Qy	903	atactaattatttgggttaaataattaatcattttttaagataattaattaagaatta	962
Db	1890		
		TAACTATAATTTGATAATCCGAATCTTAAAATGTTACAAAAATGCAAGTAAAAAATC	1949
Qy	963	aaagattttttaaanaastgataaaaatatattattcatgatgttttcatcatttgtatt	1022
Db	1950		
		ACTTATATATATATATAATATATATATATATATATATATATATATTTTAATGCACITTT	2009
Qy	1023	tfgataataaataatatttttttaatttctcaaaaaattgttgcgaacacttattagaca	1082
Db	2010		
		TAAATATAGGTTCTTTTCTTTTTTTTTTTTTTTTTTTTTTTTATGTAGAGGTGCGCAATATT	2069
Qy	1083	tagctctgtctgtttacaanaagcattcatcatttaataacattaa	1127
Db	2070		
		TTTTTTTGAATAATTAATAAAAAAGAGGTGCATATCTTCAGATTTA	2114

RESULT 10
V33135/C

ID V33135 standard; DNA; 5849 BP.
AC V33135;
DT 07-DEC-1998 (first entry)
DE Plasmodium berghei plasmid PSI-PL470 gene.
KW Malaria; infection; therapy; diagnosis; vaccine; plasmid;
KW PSI-PL470 gene; ds.
OS Plasmodium berghei ANKA strain.
PN W09835057-A1.
PN 13-AUG-1998.
PF 05-FEB-1998; IB0212.
PR 26-SEP-1997; AU-009481.
PR 06-FEB-1997; AU-004953.
PR 21-APR-1997; AU-006329.
PA (MOLE-) INST MOLECULAR & CELL BIOLOGY.
PA (UYSI-) UNIV SINGAPORE NAT.
PI Kara AKU, Nelson JS, Tan TMC, Tham JM, Ting RCY;
PI WPI; 98-447251/38.
PT Detecting Plasmodium infection from hybridisation with
PT extrachromosomal element - providing genus or species specific
PT diagnosis with few false negatives, in humans or animals
PS Claim 15; Page 54-59; 120pp; English.
CC This is the nucleotide sequence of one strand of the PSI-PL470
CC gene of the 30.7 kb extrachromosomal plasmid of Plasmodium berghei.
CC This plasmid encodes organelle-like rRNAs, tRNAs, ribosomal
CC proteins and RNA polymerase subunits, amongst others. Plasmodium is
CC detected in a human or animal sample by treating it, or derived
CC nucleic acid, with a Plasmodium extrachromosomal genetic element or
CC derived nucleic acid (A) and detecting any hybridisation. (A) can
CC include the PSI-PL470, PLH-PPH, PRB or PMQ gene, the mitochondrial
CC cox1 gene, and nucleic acids derived from them. Also new are
CC (A)-specific probes and primers (see V33139-56). The method is
CC used to diagnose Plasmodium infection. Also (not claimed) the
CC polypeptides encoded by (A) are useful as targets for drug
CC development and for development of anti-malaria vaccines. The high
CC degree of similarity between (A) from different species allows
CC development of genus- or species-specific assays that result in
CC fewer false negatives than known methods (typically 1% against 3%).
SQ Sequence 5849 BP; 2296 A; 673 C; 557 G; 2323 T;

Query Match	3.68;	Score 56.4;	DB 1;	Length 5849;
Best Local Similarity	53.13;	Pred. No. 0.23;		
Matches 120;	Conservative 0;	Mismatches 106;	Indels 0;	Gaps 0;
Qy 886	atctctctaccacatcactaattattttgggftaaataattaatcattatttttaaga	945		
Db 734	ATATTTTATATATAAATTTATATGATAAATAAATTAATAAATAAATTTTAAATA	675		
Qy 946	tattaattaaagaattaaaaagatttttttaaaaaatgtaaaaaattattattcatgat	1005		
Db 674	TTTATAAATTTTAAATATAAATATCAATATAATAATAAATAAATTTATATTAAATAGAA	615		
Qy 1006	ttttcatcacatttgattttgtataataataatatttttttaattttcttaaaaaagttgc	1065		
Db 614	ATGGATTAAATAAATAATTTGATAAATAAATTTATCTATAATAATATTATTATTTTTTA	555		
Qy 1066	aagacacttatgacatagctctgtctgtttcacaaagcattca	1111		
Db 554	TTTACAATTTTAAATATATTCTTTTAAATTTTAAATTTTAAATATGTTAA	509		

RESULT 11

Q27886/c
Q27886 standard; DNA; 2418 BP.
ID Q27886;
AC Q27886;
AD 04-FEB-1993 (first entry)
DT DT
DE DE
DE DE
KW P.falci-parum GBPl30h.
KW Polymerase chain reaction; glycophorin binding protein; inverse PCR;
KW Plasmodium falci-parum; malaria; ss.
KW Synthetic.
OS
FH Key
FH exon
FT Location/Qualifiers
FT 767..955
FT /*tag= a
FT /number= 1
FT 1111..2202
FT /*tag= b
FT /number= 2
FT 1249..2202
FT repeat_region
FT /*tag= C
PN EP-49834-A.
PD 26-AUG-1992.
PD 27-JAN-1992; 101271.
PP 21-FEB-1991; DE-105348.
PR (BEHW) BEHRINGER AG.
PR Hündt E, Knapp B, Kupper H, Nolte D, Kuepper H;
PI WPI: 92-286009/35.
PT DNA coding protein GBP 130 h and proteins produced - used for
PT vaccines to control malaria
PS Claim 1; Page 12-14; 17pp; German.
CC Glycophorin binding protein 130h is a P.falci-parum blood stage
CC antigen 69% homologous with the known GBPl30. The GBPl30h gene and
CC its homologues can be used in vaccines to protect against malaria
CC caused by P.falci-parum. The coding sequence was isolated using PCR
CC techniques on a genomic P.falci-parum FCER strain library. The
CC GBPl30h gene was found to be highly conserved between different
CC strains but is distinct from the GBPl30 gene.
CC See Q27878-Q27885 and Q33432.
SQ Sequence 2418 BP; 1047 A; 303 C; 292 G; 776 T.

	Query Match	3.6%	Score 56.2:	DB 1;	Length 2418;
	Best Local Similarity	44.3%:	Pred. No. 0.25;		
	Matches 278;	Conservative 0;	Mismatches 343;	Indels 6;	Gaps 1;
<hr/>					
Qy	520	t t t a t t c a a g c a t a t c c c t t t c a a a c t c t c t a c t t c g t c g d t t c g g g t a a t c g g	579		
Db	737	T G T T T T T C T A A T T G T G T G T T T C A A A A T A A T T A A A T A T A A A T A T A T A T A T A T A T A	678		
<hr/>					
Qy	580	t a a c a t t a g a c t t c c a a a t c a t t t t t a a c c c c t a a a c a g t a a a t t t g a a g g a c a a a a t	639		
Db	677	T A T T T T T T T T T T T T T A A T A T A T A T T A T T T A A A T A A T A A A A T A A A A A A A A A A C	618		

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 2, 2000, 02:38:02 ; Search time 5217.98 Seconds
(without alignments)
538.352 Million cell updates/sec

Title: US-09-464-528-15
Perfect score: 1574
Sequence: 1 atcgtatgagacatgttatt.....gttttgaagtataaagatg 1574

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 972840 seqs, 892348106 residues

Total number of hits satisfying chosen parameters: 1945680

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1: gb_bal.*
2: gb_ba2.*
3: gb_om.*
4: gb_ov.*
5: gb_pat.*
6: gb_ph.*
7: gb_pl1.*
8: gb_pl2.*
9: gb_pl3.*
10: gb_pr2.*
11: gb_pr3.*
12: gb_ro.*
13: gb_sts.*
14: gb_sy.*
15: gb_un.*
16: em_fun.*
17: em_hum1.*
18: em_hum2.*
19: em_in.*
20: em_om.*
21: em_or.*
22: em_ov.*
23: em_pat.*
24: em_ph.*
25: em_pl.*
26: em_ro.*
27: em_sts.*
28: em_sy.*
29: em_un.*
30: em_v1.*
31: gb_htg1.*
32: gb_htg2.*
33: gb_in1.*
34: gb_in2.*
35: em_bal.*
36: em_ba2.*
37: em_hum3.*
38: em_hum4.*
39: gb_pr4.*
40: gb_htg3.*
41: gb_htg4.*
42: gb_htg5.*
43: gb_htg6.*

44: gb_htg7.*
45: em_htg1.*
46: em_htg2.*
47: em_htg3.*
48: em_hum5.*
49: gb_pl3.*
50: gb_pr5.*
51: gb_htg8.*
52: gb_htg9.*
53: gb_htg10.*
54: gb_htg11.*
55: gb_htg12.*
56: gb_htg13.*
57: gb_htg14.*
58: gb_in3.*
59: gb_htg15.*
60: gb_htg16.*
61: gb_htg17.*
62: em_htg4.*
63: em_htg5.*
64: em_htg6.*
65: em_htg7.*
66: em_hum6.*
67: gb_htg18.*
68: gb_htg19.*
69: gb_htg20.*
70: gb_htg21.*
71: gb_htg22.*
72: gb_htg23.*
73: gb_htg24.*
74: gb_htg25.*
75: gb_htg26.*
76: gb_htg27.*
77: gb_htg28.*
78: gb_htg29.*
79: gb_htg30.*
80: gb_htg31.*
81: gb_v11.*
82: gb_v12.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	118.4	7.5	4601	33	DMU11584
2	118.4	7.5	19517	58	DMU37541
3	99.8	6.3	86829	33	PFMAL3P5
4	98.4	6.3	86829	33	PFMAL3P5
5	97.4	6.2	14867	34	AE001398
6	92.6	5.9	146285	39	AC005083
7	91.8	5.8	67970	33	PFMAL1P3
8	91.2	5.8	153098	33	PFMAL3P2
9	91.1	5.8	5371	7	YSCWTTGSA
10	89.6	5.7	67970	33	PFMAL1P3
11	89.6	5.7	104992	41	AC005504
12	89.6	5.7	130281	60	AC004157
13	89.6	5.7	161891	55	AC008206
14	88.8	5.6	253307	33	PFMAL3P7
15	88.2	5.6	80518	31	PFMAL13PA
16	88	5.6	170125	41	AC007465
17	87.4	5.6	4601	33	DMU37541
18	87.4	5.6	19517	58	DMU37541
19	86.8	5.5	104992	41	AC005504
20	86.8	5.5	130281	60	AC004157
21	86.8	5.5	152209	11	HS1108D11
22	86	5.5	15421	33	PFCOMP1RB
23	85.2	5.4	14001	33	PFCOMP1RB
24	85.2	5.4	15421	33	PFCOMP1RA
					U11584 Drosophila
					U37541 Drosophila
					AL034556 Plasmodiu
					AL034556 Plasmodiu
					AE001398 Plasmodiu
					AC005083 Homo sapi
					AL031746 Plasmodiu
					AL034558 Plasmodiu
					J01462 Yeast (S.ce
					AL031746 Plasmodiu
					AC005504 Plasmodiu
					AC004157 Plasmodiu
					AC008206 Drosophil
					AL034559 Plasmodiu
					AL109815 Plasmodiu
					AC007465 Homo sapi
					U11584 Drosophila
					U37541 Drosophila
					AC005504 Plasmodiu
					AL034559 Plasmodiu
					AL034419 Human DNA
					X95275 P.falciparu
					X95276 P.falciparu
					X95275 P.falciparu

```

25      85      5.4 161286 67 AC025120
c 26      85      5.4 173315 9 AP001254 Homo sapi
27      84.8      5.4 18977 8 U32857 Saccharomyc
c 28      84.8      5.4 3542 7 YSCMTG016
29      84.8      5.4 157781 55 AC011355
30      84.8      5.4 175516 60 AC006280
c 31      84.4      5.4 34996 34 U87145
c 32      83.6      5.3 163678 52 AC010178
c 33      82      5.2 6591 7 YSCMTG006
c 34      81.6      5.2 2226 33 MIDVTRN
c 35      81.6      5.2 75076 39 AC004948
c 36      81.2      5.2 152209 11 H51108D11
c 37      81      5.1 14867 34 AE001398
c 38      80.8      5.1 1400 34 AF151382
c 39      80.4      5.1 85779 8 SC0011856
c 40      80.4      5.1 146285 39 AC005083
c 41      80.2      5.1 110000 31 PEMAL4P1_0
c 42      80      5.1 94384 39 AC011718
c 43      79.6      5.1 161891 55 AC008206
c 44      79.6      5.1 193561 39 AC004801
c 45      79.4      5.0 170109 39 AC008079

```

ALIGNMENTS

```

RESULT 1
DMU11584
LOCUS      DMU11584      4601 bp      DNA      INV      23-JUL-1994
DEFINITION Drosophila melanogaster Oregon-R mitochondrial A+T region.
ACCESSION  U11584
VERSION     U11584.1 GI:508826
KEYWORDS   mitochondrial DNA; A+T region; tandem repeats.
SOURCE     fruit fly.
ORGANISM   Mitochondrion Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE  1 (bases 1 to 4601)
            Lewis,D.L., Farr,C.L., Farquhar,A.L. and Kaguni,L.S.
            Sequence, Organization and Evolution of the A+T Region of
            Drosophila melanogaster Mitochondrial DNA
            Mol. Biol. Evol. 11, 523-538 (1994)
94265822
REFERENCE  2 (bases 1 to 4601)
            Kaguni,L.S.
            Direct Submission
            Submitted (28-JUN-1994) Laurie S. Kaguni Ph.D, Dept. of
            Biochemistry, Michigan State University, East Lansing, MI,
            48824-1318, USA
FEATURES             Location/Qualifiers
     source           1..4601
                     /organism="Drosophila melanogaster"
                     /organella="mitochondrion"
                     /strain="Oregon-R"
                     /db_xref="taxon:7227"
                     /dev_stage="embryo"
     gene             1..4601
                     /gene="mt:ori"
                     /note="Mitochondrial origin"
                     /allele=""
     repeat_unit      650..1022
                     /db_xref="FlyBase:FBgn0013687"
                     /gene="mt:ori"
                     /note="repeat I-A"
                     /db_xref="FlyBase:FBgn0013687"
                     /rpt_type=tandem
     repeat_unit      1023..1360
                     /gene="mt:ori"
                     /note="repeat I-B1"
                     /db_xref="FlyBase:FBgn0013687"
                     /rpt_type=tandem
     repeat_unit      1361..1705

```

```

/gene="mt:ori"
/db_xref="repeat I-C/A"
/rpt_type=tandem
1706..2043
/gene="mt:ori"
/db_xref="repeat I-B2"
/rpt_type=tandem
2044..2388
/gene="mt:ori"
/db_xref="FlyBase:FBgn0013687"
/rpt_type=tandem
2491..2511
/gene="mt:ori"
/db_xref="FlyBase:FBgn0013687"
/rpt_type=tandem
2512..2648
/partial
/gene="mt:ori"
/db_xref="FlyBase:FBgn0013687"
/rpt_type=tandem
2649..3112
/gene="mt:ori"
/db_xref="repeat II-A"
/rpt_type=tandem
3113..3576
/gene="mt:ori"
/db_xref="repeat II-B1"
/rpt_type=tandem
3577..4040
/gene="mt:ori"
/db_xref="repeat II-B2"
/rpt_type=tandem
4041..4504
/gene="mt:ori"
/db_xref="FlyBase:FBgn0013687"
/rpt_type=tandem
complement(4565..4585)
/note="deoxythymidylate stretch"
BASE COUNT  2271 a 131 c 74 g 2125 t
ORIGIN

```

```

Query Match      7.5%; Score 118.4; DB 33; Length 4601;
Best Local Similarity 48.4%; Pred. NO. 1.2e-06;
Matches 520; Conservative 0; Mismatches 531; Indels 23; Gaps 6;

Oy  104 acttaataaaattttacatttaattagaatttttttttatacaataataattattata 163
      ||||| ||||| || ||||| || ||||| || ||||| || ||||| || ||||| ||
Db  2187 ACTTAAATAAATATTTTATTTTAAATAAATAATTTATTTATTAATTAATTAATTA 2246

Oy  164 gttttattagaaa-tattaattagaaaaattttgaatcccgatttctctctctcttc 222
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
Db  2247 TTTTATTAATAAATATTTTATTAATAAATAATTTATTAATAAATAATTAATAATA 2306

Oy  223 gctattcatcttttttaacccaacaaatcttatgttcttcaaatagaaactgaaat 282
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
Db  2307 TATATATATATATATATATTAATAAGAAATAATTTTAAATTTTAAATAAATAAAT 2366

Oy  283 tattaattataataaactgaaacaaatttggtatcaattcatatatacatgtttagata 342
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
Db  2367 TAATAAATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 2423

Oy  343 aaagcgataattatgataaaactgcaaaagattttacaaatattcagaaaaaaat 402
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
Db  2424 AATTTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 2483

```



```

171..239
/gene="mt:ND6"
/db_xref="FlyBase:FBgn0013685"
/transl_table=5
/product="ATPase 8"
/protein_id="AAC47814.1"
/db_xref="GI:1166533"
/transl_table="MPQAPISWLLLIIFISITFILFCSINYSYMPNPKSNELKNI
NLNSMNWKW"
4062..4736
/gene="mt:ND6"
/codon_start=1
/db_xref="FlyBase:FBgn0013685"
/transl_table=5
/product="ATPase 6"
/protein_id="AAC47815.1"
/db_xref="GI:1166534"
/transl_table="MNTNLSVDFDPLAFNFSNLWLSLFLGLMIPSIYWLMPNRYNI
MNSLTLTHKEFKTLGPGSGSTFIISLPSLILFNNMGLFPYIFTSRSLTTLT
LSLAFLWLCFMYLGYNHHTOHMAFLVPOGTAILMPFMCIEISNIIRPCTLAVR
LTANMIAGHLTLTLGNTGSSMSYMLTFLLMAQIALLVLESAYAMIQSYVAVLSTL
YSSEVN"
4736..5524
/gene="mt:ND6"
/codon_start=1
/db_xref="FlyBase:FBgn0013685"
/transl_table=5
/product="cytochrome c oxidase subunit III"
/protein_id="AAC47816.1"
/db_xref="GI:1166535"
/transl_table="MSTHNSHPFLVDYSPWPLTGAIGAMTIVSGMKWFHYDLSLF
VLGNIITILTYQWRDVSRECTYQGLHYAVTIGLRWGMILFISELVFFYSPFWAF
FHSLSPAELGASNPWGIIISFNFPQIPLLNLAILLASGAVTVTAHSHLMENHSOT
TOGLEFTVLLGIYFTILOAYEIEAPFTIADSIYSGTFMATGFGFHHVLIIGTITFLLY
CLLRHLNHSKHHFGEPEAAWYHWFVDVWVLFYITIIYMWGG"
5543..5607
/gene="mt:ND6"

CDS
Query Match 7.5%; Score 118.4; DB 58; Length 19517;
Best Local Similarity 48.4%; Pred. No. 8.2e-07;
Matches 520; Conservative 0; Mismatches 531; Indels 23; Gaps 6;

QY 104 actaataaaattttacattaaatttagaattttttttacataataataatttatta 163
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 17103 ACTATAAAATAATTTTTTTTTAAAAAAATAATTTATTAATAATTACTTAATAACTA 17162

QY 164 gttttattagaaa-tattaattagaaaattttgaatcccgatttcttccctttcttc 222
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 17163 TTTTATAATAAATTTATTTATAATAAATAATTTATAAATAATAATAAATAATTTATA 17222

QY 223 gctattcatcttttctaaacaaacatttataatgtcttccaaattagaaattgaaat 282
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 17223 TATATATATATATATATATAAATAAATAAATAAATAAATAAATAAATAAATAAAT 17282

QY 283 tattaattataataacgaacaaatttggtatcaattcatatcatagcttagtaata 342
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 17283 TAATAATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAAT 17339

QY 343 aaatgcgataaattgaataactctgcaaaaagattttacataatactttccagaaaaat 402
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 17340 AAATTTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAA 17399

QY 403 taataacaaatttgcgttttctcatggtggtgagagagatttgcgcactatagaac 462
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 17400 AAAAATTAATTTTTTTTTTTTTTTTTTAATAATAATAATAATAATAATAATAATAA 17459

QY 463 tctccacagaccattcttgcactcaacaaacattggtcagaattggtggggatttt 522
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 17460 TTTATATATATTGAATATTTATATAATATATATATATATATATAGAAAAATTAATTTAA 17519

QY 523 atattcaagcataccctttcaaaacttctacttacttctcgctcgctcgtaacggtaa 582
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 17520 ATAAATTAATAATAATTTTTTAAAAAATTTCT--TAAATGTATTATTTTTTATAAAAAATAT 17577

```



```
Oy 661 actatttttttttgaatttcacgaacaaacacagatttatcctgaatttttagaacc 720
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 39502 ATAAATTTATTTTATTTTAAATAAATAATATATATTAATTTATTTTAAATAAATA 39443

Oy 721 acagatgtaacaaacaaatattattttttctaaacaaatttcacgcagcagtc 780
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 39442 ATATATTAATTAATTTATTTTAAATAAATAATATATTAATTTATTTTAAATA 39383

Oy 781 ctgagccatgaaaaaacctcataaaaatactcacacattgaccattgaaaagttcgtt 840
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 39382 TAAATAATATATATTAATTTATTTTAAATAAATAATATATTAATTTATTTTAA 39323

Oy 841 ctcccatggtgtaacagatcaaacctcacacataacatacattggtgattcctctacca 900
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 39322 TTTTAAATAAATAATATATTAATTTATTTT-----TATTTTAAATAAATAATA 39272

Oy 901 tcatcctaattatttgggttaataattatcattattttttaaagataatttaataagaat 960
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 39271 TTAATTTATTTTAAATAAATAATATATTAATTTATTTTAAATAAATAATA 39212

Oy 961 taaagattttttaaanaaatgataaaaattattatttattcatgatttttcacatttga 1020
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 39211 TATTTATTTATTTAAATTAAT--TTAAATAAATAATTTATTTTATTTTATTTAT 39154

Oy 1021 ttttgataaataattttttttaaatttcttaaaaaatgttgcaagacactattaga 1080
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 39153 TTAATATATATATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 39094

Oy 1081 catagcttctgtctgttcaaaaagcattcatcatttaatacattaaaaattttaata 1140
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 39093 TATATATACTATTACATATATATTTTATTTTATTTTATTTTAAATAATTTATTTAT 39034

Oy 1141 ctaacagtagaatctctctgtgagtt 1165
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 39033 TAAATAATAAATAATTAATTTATTTAT 39009

RESULT 4
PFMAL3P5 86829 bp DNA INV 11-FEB-2000
LOCUS Plasmodium falciparum MAL3P5, complete sequence.
DEFINITION AL034556 AL008971 AL008972 AL008978 AL010141 AL010153 AL010162
ACCESSION AL010206 AL010210 AL139179
VERSION AL034556.2 GI:4493931
KEYWORDS HTG.
SOURCE malaria parasite P. falciparum.
ORGANISM Plasmodium falciparum
REFERENCE 1 (bases 1 to 86829)
AUTHORS Bowman, S., Lawson, D., Basham, D., Brown, D., Davies, R.M., Devlin, K., Fellows, T., Churcher, C.M., Craig, A., Hamlin, N., Harris, D., Holroyd, S., Hornsby, T., Horrocks, P., Jagsal, B., Jagsal, B., Kyes, S., McLean, J., Mould, S., Mungall, K., Murphy, L., Oliver, K., Quail, M.A., Rajandream, M.-A., Rutter, S., Skelton, J., Squares, R., Sulston, J.E., Whitehead, S., Woodward, J.R., Newbold, C. and Barrell, B.G.
TITLE The complete nucleotide sequence of chromosome 3 of Plasmodium falciparum
JOURNAL Nature 400 (6744), 532-538 (1999)
MEDLINE 99376085
REFERENCE 2 (bases 1 to 86829)
AUTHORS Bowman, S., Skelton, J., Churcher, C., Lawson, D., Quail, M. and Barrell, B.
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 86829)
AUTHORS Lawson, D., Bowman, S. and Barrell, B.
TITLE Direct Submission
JOURNAL Submitted (17-DEC-1998) P. falciparum Genome Sequencing Consortium, The Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK
COMMENT On Mar 24, 1999 this sequence version replaced gi:4034877.
For more information about this sequence or the Malaria Project,
```

```
see http://www.sanger.ac.uk/Projects/P_falciparum.
Location/Qualifiers
1. 86829
/organism="Plasmodium falciparum"
/strain="3D7"
/db_xref="taxon:5833"
/chromosome="3"
/clone="MAL3P5"
324. .2944
/gene="MAL3P5.1"
join(324..668,1199..1303,1460..2944)
/gene="MAL3P5.1"
/note="predicted using hexExon; MAL3P5.1 (PFC0575w),
Hypothetical protein, len: 645 aa"
/codon_start=1
/protein_id="CAB38969.1"
/db_xref="GI:4493933"
/db_xref="SPTREMBL:O97258"
/translation="MYLKNVYIYISCFILDFCLFSLHLLKMKYKNNHNNMKSVTFFEL
RSPQYRKRFRKRIRKYNVSPFKKKQKPLFLFENLKKGFSLGFWNQYDQVKNRKKK
KKKKKKKKPKVHSILNOISEKVKKKDAENYALHLFLKLDENITLFSNMHIMDF
FKSKQVIECDIRDKSKRRKKNLSIYNLFCTLIYTYCHCLIKVISHLICIFEFF
FFCFCLYNILRIYEECVGDIIRKKIERINYSNDINNDISIDKDMKNNPDIIVNTNI
LYFNHYDELRCFTMKLNERNKNIIRSNYDINNDISIDKDMKNNPDIIVNTNI
NNNTMKKKHKIKDTYDDDDYDEKEDLVQKNDIYKIEDQPLYNNDSNENINDNN
SLDEKIKQFNPDDENELKDYEQFLPNDNIYKIEDQPLYNNDSNENINDNN
TEADIEFQNFSLNVLQOHGRVKSNDENTKSTHINKNTINKGYDTLQNLQENNN
FKKNDNNISNDISNEINIILKLNQSDENILNLSDIYERLRTKVLNTYQKIEY
LKFKYQYDIINQDPIIKNEKTVLDLLNYGKIVMSPDVNSLFEKTKIDSLPNEKDK
NNQENQKSNKYIIMFNM"
complement(3356..6646)
/gene="MAL3P5.2"
complement(3356..6646)
/gene="MAL3P5.2"
/note="predicted using hexExon; MAL3P5.2 (PFC0580c),
Hypothetical protein, len: 1097 aa"
/codon_start=1
/protein_id="CAB38968.1"
/db_xref="GI:4493932"
/db_xref="SPTREMBL:O97257"
/translation="MLGUKRNVFYLLVSVLSFAYFLKRRKDNENNYETLINNDIE
KRRRIHNSKSYIPLEFLNIYDSYIKNKLRLWLYFKPKRKRKDEEYIITNMVRK
KREAIKYNFISDEQNLNFYIYVLEYSKYLGLSPHLSLYILKNISEHCVNIYP
SLYYNKLNDKHLNKKELKFKQIINNEHTQAAPTNTHTNNNNNNKKPLDINHSC
KNTNYSSTYNNMEKENIYDKYNIHNFYTESISYKDNCOHTLNMVILLNQTY
DNICRLTNTNTIYNFMNLKAYICYKNMEILLAYNHIEDMKKIKNKNNTS
FYIYISFFFKKEENHIYDFEDQMHHLKENDKPKFNSNETHNNIYKISONY
LFDHINSNRCSFKNLKKQDDNTKTHIMGKPKMNSDHNKNNNTCNINEX
DOKDILAKIYFLKGLDDIOILNELYVIMYRLLFESCLISIKKNIHLKKME
PDKDKIYLSADYNNLRNLKRFKNEERENINSFASPLLSKNIYFEDEIG
RSDNTIYNNVYDKETNTTNNNDNNNDNCSNNDHICSNNNDHCSNNNDHICSN
NNNCSNNNNNCSNNNNNCSNNNDNCSNNNDHICSNNNDHCSNNNDHICSNN
YELAKIYNNIFDLKGRKQNDNTNLIYIKKRPWILFKNIKNKDTSFIEH
NNNIDKNNNIIFKKYNNFESSIYFYIYKDYENKRLRYIYDNIKKKFR
FVKWNERIKRLKMFRAHYIYFDOFIINNYHIIHKKNIHIIHKKQCKDE
IDIVFKDLYCMINNIINNIYFVYIKVDHNECVRYEKAYNKILLYEYNLAKENIY
YANKTKLYTLNNISNDLYPNISYNNKIYQNKYKKNKNSHFIYSKLNKDIHLLE
LYTQIQNCCDIFSYIKYFNKFNKPNFLNLYELHUYIYSEKKKKKFFSFISSP
YSYDTMVSFTSFYFFSLSYLLFLYHPDMYASYIFFEKTUTYSGLPYIYVLYNNI
MVCGPKT"
complement(13655..19355)
/gene="MAL3P5.4"
complement(join(13655..17801,18019..19355))
/gene="MAL3P5.4"
/note="predicted using hexExon; MAL3P5.4 (PFC0590c),
Hypothetical protein, len: 1828 aa"
/codon_start=1
/protein_id="CAB38986.1"
/db_xref="GI:4493950"
/db_xref="SPTREMBL:O97275"
/translation="MLFYKLCILFLSFIHIIHGKDISSTKLLKSGNLNNKIKKGR
RYGGIYNSKMYTIEKRRYIKKQKNNNSYFCNVYKNNDVNNNTAYNHYIYNNP
KNKLEYEKIKNHVKKKKKIFSLKFSQNKREKKKKYFFINFTSFHDIKDNKIVLF
```



```

QY 587 agactttcaaaatcatttttaacccctaaacagtaaatgttgaggacaaataatattt 646
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 7863 AAAACATATTATA---TTTAAATATATAAAATGCTATATATAAAATAATTTCTTTAA 7918
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 647 tcaaatgttagagactatttttttttgaatttgacgaacaaacacagattatcct 706
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 7919 TTTTATTTTATTAATAATTTTATTTTATTTTATTTTATTTATATATATATATAT 7978
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 707 gaatttttagaacacagatgtacaaacccaattattttttttctaaacaaaatt 766
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 7979 TTTTAAATGTTAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 8038
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 767 tcatgacagatccctcagccctgaaacaaacccctataaaaaatactcacacattgacca 826
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 8039 ATAATTATATATATATATATATTTTACATATAAAATAATATATA-----TTATATATATTTAA 8093
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 827 ttgaaagttcgttcctccatcggtgaacagatcaaacacacacacataacatacgg- 885
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 8094 TTTTATTTTATTTTATTTTAAATGAATAAATAAATAAATAAATAAATAAATAAATAA 8153
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 886 atactcctccacacatacactaattattttgggtggttaaatatttaacatttttttaaga 945
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 8154 AATATTATTTATATATATTTAAATTAATAATTAATAATATATTTATTTTAACTTAATTA 8213
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 946 tataattaagaataataaagatttttttaaaaaaaagtataaaatbatattatcattgat 1005
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 8214 ATTAATATTTATTTTAAAAAAAATAATATAAATATGATATATATATAATATAATATA 8273
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1006 ttctacacatttgatttgataaaaaataatttttttttaatttttttttaaaaaatttgc 1065
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 8274 TATTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 8333
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1066 aagacactattagacatagctgtctgtttacaaagcattcatcatttaacatt 1125
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 8334 ATATTATTTATGACTCTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 8393
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1126 aaaaaatattt 1136
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 8394 TAATAATATAT 8404
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

RESULT 6
AC005083/c 146285 bp DNA PRI 21-DEC-1999
LOCUS Homo sapiens BAC clone CTA-281G5 from 7p15-p21, complete sequence.
DEFINITION AC005083
ACCESSION AC005083
VERSION AC005083.1 GI:4150930
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 146285)
AUTHORS Madsen,C. and Blair,T.
TITLE The sequence of Homo sapiens BAC clone CTA-281G5
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 146285)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (12-JUN-1998) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 3 (bases 1 to 146285)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (12-JAN-1999) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE 4 (bases 1 to 146285)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (21-DEC-1999) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT On Jan 12, 1999 this sequence version replaced gi:3212908.

```

```

----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: sapiens@wustl.wustl.edu
----- Summary Statistics
-----
Center project name: H_RG281G05
-----

```

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded or sequenced with an alternate chemistry; an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/CTB/CHR7>, send <mailto:egreen@nhgri.nih.gov>, or see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

Clone CTA-281G5 is from a release of the human BAC library CITB-HS-A. The library contains cloned DNA from human sperm. See: Shizuya et al., Proc. Natl. Acad. Sci. USA 89:8794-7 (1992); U-J. Kim et al., Genomics 34:213-8 (1996). The clone is available from Research Genetics, Inc. (<http://www.resgen.com>). VECTOR: pBelOBAC11 Selection: chloramphenicol

NEIGHBORING SEQUENCE INFORMATION:

Actual start of this clone is at base position 1 of CTA-281G5; actual end is at 146285 of CTA-281G5

BAC CTA-281G5 contains an E. coli transposon from 10982 to 12345 that is not represented in the submitted sequence.

FEATURES	source
repeat_region	1..146285
repeat_region	/organism="Homo sapiens"
repeat_region	/db_xref="taxon:9606"
repeat_region	/chromosome="7"
repeat_region	/map="7p15-p21"
repeat_region	/clone="CTA-281G5"
repeat_region	/clone_lib="CITB-HS-A"
repeat_region	322..358
repeat_region	/rpt_family="TAAAA)n"
repeat_region	808..952
repeat_region	/rpt_family="MER1_type"
repeat_region	1172..1194
repeat_region	/rpt_family="AT_rich"
repeat_region	1264..1720
repeat_region	/rpt_family="LI"
repeat_region	1848..1881
repeat_region	/rpt_family="AT_rich"
repeat_region	1882..2171
repeat_region	/rpt_family="Alu"
repeat_region	3782..3804
repeat_region	/rpt_family="AT_rich"
repeat_region	4794..4826
repeat_region	/rpt_family="AT_rich"
repeat_region	5764..6034
repeat_region	/rpt_family="purine-rich"
repeat_region	6891..7126

KKIPLVAGTYKYIDEPSLNKINNYALKNOKIGIVGSGAGKSTILLSILGLINISOG
KITVEGRDITRYNRKGDSDSIIIGIAQSSFFVFNWNIIRTFIDPYNFNTDDEIVHALKIN
GINLGNLDLYKXHHKQDKSNYKIIOTSKVINSNDNTILLNDICIRLSYLRPLYLN
RHKTKIILIDRIPFNUNNSVHDLGNSFLGKAKSFNYIIRNHPFNNTVLIISHHANT
LSCDDYIYLRKGBITRYCSYEDVKTOSELSHLEMD0"
23896..31533
/gene="rRNA"
/note="region containing small subunit, 5.8S and large
subunit rRNA genes and spacer regions"
23896..31533
/gene="rRNA"
complement(31966..32775)
/gene="MALP3.04"
complement(join(31966..32476,32675..32775))
/gene="MALP3.04"
len: 203 aa, similarity: P. falciparum chromosome 2,
PF0110W, O96126 predicted integral membrane protein (255
aa), fasta scores: opt: 335, E(): 4.9e-15, (36.1% identity
in 191 aa overlap)
/codon_start=1
/product="conserved hypothetical membrane protein,
MALP3.04"
/protein_id="CAB63559.1"
/db_xref="GI:6594247"
/translation="MKKSYTFINVTILLFLTLFLTYNYDYFSKTKFNNNIKIDIN
RFKRIAEASEOKYPWEEDPCLTNEELIRPEHNDSPVLPHEHYENIDKINELISINS.
TKTKWETIKMRQNYEKETDMNHNWDFMWHYKWKANLYLYKVKHLINILKDLTNP
HDKETITTIWIKQIEDIEYFLNQLQWELRIULTLELFYNKE"
complement(32477..32486)
/gene="MALP3.04"
/note="potential splice acceptor sequence"
complement(32669..32674)
/gene="MALP3.04"
/note="potential splice donor sequence, aaa/gtatat"
36657..37343
/gene="MALP3.05"
join(36657..36743,36864..37343)
/gene="MALP3.05"
/note="MALP3.05, hypothetical protein, len: 188 aa"
/codon_start=1
/product="hypothetical protein, MALP3.05"
/protein_id="CAB63560.1"
/db_xref="GI:6594248"
/translation="MRKMSGIFFIKLLICISFICFCFNKCMISYRKDLLWYSEN
CFNYSIDRSLAEGSESSEKTKVDIPNIELLSLNINVEYKMKETVGFPMNNLIN
IANVLIHISFNINENIFSLINDSSKSPVLKFLKEFGSIFPHMLNNVPKLLFDLCQ
RNPVHILGLIVILAAIYVFENFKNEC"
36744..36749
/gene="MALP3.05"
/note="potential splice donor sequence, aag/gtatga"
36854..36863
/gene="MALP3.05"
/note="potential splice acceptor sequence"
complement(38049..40284)
/gene="garp"
complement(join(38049..39995,40210..40284))
/gene="garp"
/note="MALP3.06, garp, len: 673 aa, similarity: almost
identical to GARP_PLAFF (678 aa), fasta scores: 97.6%
identity in 678 aa overlap"
/codon_start=1
/product="hypothetical garp protein"
/protein_id="CAB63561.1"
/db_xref="GI:6594249"
/translation="MNVLFSLNYICILFFVVCITLNFSTKCFNSGLLNKINLKSFSOS
ITGLLNETLEKNKNDKSDSETLLKEEKDEKDDVPTTSDNLKNAANNNEISSSTOP
TNIINVDKONENSVDKKDKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK
DKKKEENSVSLYTKQHPKPKNATEHGENLYEENVSEINNAQGGILLSSPYQRE
QGGCGIISSEVHNTNDTKDENISDKKEDHQOHEMLTLDKKERKQKEKKEQ
KIEKKKKKQKEEKKEKKEKKEKKEKKEKKEKKEKKEKKEKKEKKEKKEKKEK
HDKNEETMOQDQTSSETNNEINVLPSPLTDVTTPEEHKEGHEEKEGHEGHEG
EHKEEHEEKEEHEKKEHKSCKGKDKGKKKKKKKKKKKKKKKKKKKKKKKKKK
VIEDE

DKGVEIINLEDKACEQHTVBSRPLSQPOCKLIDPEOLTLMDSKVEEKNLSIQ
EOLIGTIGRYNVVPRRDNHKKMAKIEAELOKOKHYDKEDKKEEKEEKEEKEVO
EDDEVEDE
DDDE
complement(39996..40005)
/gene="garp"
/note="potential splice acceptor sequence"
complement(40204..40209)
/gene="garp"
/note="potential splice donor sequence, aag/gtaaca"
45401..50233
/gene="MALP3.07"
Query Match 5.8%; Score 91.8; DB 33; Length 67970;
Best Local Similarity 46.3%; Pred. NO. 0.0012;
Matches 487; Conservative 0; Mismatches 547; Indels 17; Gaps 5;
Oy 92 aaaaatcgagacttaataaaattttacatttaaatagaaattttttttatcaataaat 151
Db 8514 AATATATAATTAATAATAATAATAATTAATAATAATAATAATAATAATAATAAT 8573
Oy 152 attaatatttagttttattagaaatttaattagaaattttgaatcccgatttctcc 211
Db 8574 ATAAATTAAAT 8627
Oy 212 tcttttctcgctattcatctatttcttaacaaacacacatttcttcttcaaat 271
Db 8628 ATAATATAATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 8687
Oy 272 gaacttgaatttaatttaataataactgaacacacatttgggtcaattcatatacat 331
Db 8688 TAAATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 8747
Oy 332 gcttagtaataaaatcgataatttaattgataaaatctgcacaaagattttacaaatctt 391
Db 8748 GTTAAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 8807
Oy 392 tcagaaaaataataacaaatttgcgtttctgcgttgcgtgcgtgcgtgcgtgcgtgcgt 451
Db 8808 ATTATATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 8867
Oy 452 cactatagaactctctcgcgcacatttcttgcacttcaactcaacacgattgctcgaattg 511
Db 8868 TAAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 8926
Oy 512 gtgggatttattatcagacatcccccttcaaaccttccacttacttctgcgtgcgttgc 571
Db 8927 TAAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 8986
Oy 572 gtaatcggtaacatttagactttcaaaatcatttttcaacccctcaaacagataatttgaagg 631
Db 8987 TTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 9044
Oy 632 acaaaaaataatttttcaaaatttgatgactatttttttttttttttttttttttttttttt 691
Db 9045 TTTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 9104
Oy 692 aacagattattctctgaatttttaggaacacacagatgtaactaaacacattttattttatt 751
Db 9105 AATTTATATATATAATAATAATAATAATAATAATAATAATAATAATAATAATA 9164
Oy 752 ttctaaaaacaaatttcattgcgcagcattccctcagcccatgaaaaaaccttataaaaaa 811
Db 9165 TTAATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 9224
Oy 812 tctcacacattgacattgaaagattcgttctctccattgggtacaccagatcaaacatcacatc 871
Db 9225 TATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 9284
Oy 872 caaacataacatggatctctcttaccatcatacataactatttttgggttaaatattat 931
Db 9285 TTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 9338

COMMENT

On Dec 16, 1999 this sequence version replaced gi:5763807.
 For more information about this sequence or the Malaria project,
 see <http://www.sanger.ac.uk/Projects/P-falciparum>. IMPORTANT: This
 sequence is unfinished and does not necessarily represent the
 correct sequence. Work on the sequence is in progress and the
 release of this data is based on the understanding that the
 sequence may change as work continues. The sequence may be
 contaminated with foreign sequence from E.coli, yeast, vector,
 phase etc.

FEATURES

```

source      Location/Qualifiers
1..67970
  /organism="Plasmodium falciparum"
  /strain="3D7"
  /db_xref="taxon:5833"
  /chromosome="11"
  /complement(1748..3276)
  /gene="MAL1P3.01"
  /complement(1748..3276)
  /gene="MAL1P3.01"
  /complement(1748..3276)
  /note="MAL1P3.01, conserved hypothetical protein, len: 412
  aa, similarity: upf0006 family eg to
  YBL055C/YBL0512/YBL0511, YBF5_YEAST (418 aa), fasta
  scores: opt: 316, E(): 1.1e-12, (33.2% identity in 271 aa
  overlap)"
  /codon_start=1
  /product="conserved hypothetical protein, UPF0006 family"
  /protein_id="CAB63556.1"
  /db_xref="GI:5594244"
  /translation="MKLVFHYIKYINLVFYSIIPLKSNLSUKIYNDLRYISTVKNKYV
  LQKKRLNKHNRIMKEDNESSFIDGSLNLTDMFVGYNKSKHENDLQNLNRAK
  NNKKRIITCTCLAEIDKSLKICETDPEGKFLYLSAGVPTNCYEFIDKNKHEEKE
  IIAKVEEFTKYFKNQVENSWMGNGKKICDGEKDMNNEILLENKLDITPEGKY
  NEKDEVELNKKLIKPNPVICIGIETGIDPDRLYPCSKYIQIKYFIQKLQVGMN
  LPMELHRNGSETFEKVIDYKFLERNGVGHSTEDIVHIVQNKRLYGVNG
  CSLKLEINAVKPIPLNLLLETADPCWGVKKTHASVEYIKDTYKRAYTNLKKIKN
  IIKDDNTIFKREPNYNIA"
  /complement(2599..2610)
  /gene="MAL1P3.01"
  /note="potential splice acceptor sequence"
  /complement(2742..2747)
  /gene="MAL1P3.01"
  /note="potential splice donor sequence, atg/gttaa"
  /complement(2849..2861)
  /gene="MAL1P3.01"
  /note="potential splice acceptor sequence"
  /complement(2984..2989)
  /gene="MAL1P3.01"
  /note="potential splice donor sequence, aaa/gtaaaa"
  /complement(5005..5496)
  /gene="MAL1P3.02"
  /complement(5005..5496)
  /gene="MAL1P3.02"
  /note="MAL1P3.02"
  /gene="MAL1P3.02, hypothetical protein, len: 163 aa,
  contains possible signal sequence"
  /codon_start=1
  /product="hypothetical protein, MAL1P3.02"
  /protein_id="CAB63557.1"
  /db_xref="GI:5594245"
  /translation="MKLLNRFVYVLCPIILFFELNSVLGNRRNNINFHETNAAK
  AMRKLGSINSLKDRGDELKILNDEKHKDSTKWKSYSFISNLEEKYSOTDLFR
  KKOINEANTKIIEDRQEFILNDEIENIATRFVLENNFDELVIQSKQLSDILOS
  LNN"
  /complement(8020..10389)
  /note="possible cen1, region of very high [A+T] content"
  /complement(14884..20352)
  /gene="MAL1P3.03"
  /complement(14884..20352)
  /gene="MAL1P3.03"
  /note="MAL1P3.03"
  /complement(14884..20352)
  /gene="MAL1P3.03"
  /note="MAL1P3.03, putative ABC transporter, len: 1822 aa"
  /codon_start=1
  /product="putative ABC transporter"
  /protein_id="CAB63558.1"
  /db_xref="GI:5594246"
  /translation="MTTYKENVGISNKGKKKKSCQNISFLNLSFDWIRPLINDLIK
  GDIELPNICRNPDPVYASKLEENLROIEVEDSEFYSEKNSNEHVLHHCNSDASE
  KKVYVYHYHNLWSILATKFRILIIISFYILETLVTLGGKFDYDMYRILEGOKIPV
  YISLDFKVPFSGLVVYVNMFFLFFALLHFFHFLHTINLKSVMFLYKINLCSSNN
  NHLQNPAYNTYTRFSQTEIDISDRDLSIGKNASSSSSGKNNKNNKNDKNNKNN
  DIINFIKSTKMEKSLNENRSLPNYINMFSDVPSVTFYFYSINLPNFKVFIK
  MSFYVFIKIGSNSVGIAIWLISALYSAMILFELPSLFKSLYLYRDKRDNNHHVL
  KFLKLMFNWSEFAFYINIFRMKMKYKIRLYLSNIGVFISSIDIVVYVFI
  YKDLANKKEIKETSIIPLYVVKILISNVANFNPLVNVNMGIVNIKRLNVIINDH
  LYNDIKNYFMYRYNEDYNIYVDKTFLONETISDHDGTSNHLKHLKNVYKNNLTN
  MFKYFFTHKNNYHNKNIILQSLGLKNDVDDNKKICFQEHKSNSTYNTNSHIHE
  KKEYENIHNSNSTMSEKFKKNNKNEYIILKENCDFGLSDYDKNDCHILKNINFN
  LKNSMLAIIGNVGSGSAFPHISLQFDMTHGNLYIENFFKMPILYVPPNSWLFMG
  NIRSMLGNEYNPLYIKYTIQSELNDLSTIEHGMKYINDDNHLSKGVKVICLA
  RALYEHVTHHKLCTDYKELIOPNEILDKDLNNKNTSSYNNKSKLVNYPNEN
  YLOKCLMDDNNFYLYLLDDIFSDPSISKKIFSNLCKEDNIFSKNCSPILSMKS
  TDLNLEDILDNVOYEIVNFEIOQKLTIRGNISEYMEKNNLNTKESHGWSNLT
  IDYTRIKLDEVELNHVKHNSKMIYKAYFVKNGTESVSFSEIDSEIKYIKMKKKNY
  KKEHNNKNNKNNNNNNKDDHINIMNDNHRNYNDINLGNSTDDSPSTVSLGNE
  YTLDTYTSNNSDKDEIVKPLKYTHHEFNKSSMPFVKSSNMNPNFKNKESHS
  FKGSISLETLYWYFOOVGVLLTSVIFMLISIFTDEIKFVFLTMSTISKNKESHS
  TILQKQVRYLEYFVILPIISLVTSGICESMIYIGNITSAIKVHNHILYSLNAPLYP
  YNNLGNLIRFIIDISAFDYGFLKRIKAFIFRCLISLLIYIWRDCIFIPFPV
  ILIYFFVFRKFSRGCKEAQRLYLSCHTPLCNIYSNALSGKNIINIKKNTYHLDVYE
  HYNNFRISYFFKWLINWASLYIKIFILLTTYIMHPLYASGIKLYKRYKVR
  LSTGYCISFSARGVITKFLCDYTHIEKEMCCVQRLPEPAKISNKNASMMKNEL
  NVITQYKKNENISDKISAIVEYKNVSLSIINSQDDEKKYGGIKPENYVSYK
  KKITLVNGTYKIDEEPGLKINMYALKNOKIGIVGSGACKSTILLISLNLISQC
  KITVEGDIRTYNRKGEISIIQIQAQSFVFNWNTIRTFIDYNNFTDDEILVHAKLN
  GINGLNDLYKMHKQDKMSYKKIQTTSKYINSQNTILLTNDICIRLSVLRLYN
  RHKYKILIDIEIPFNLSVHDELNGLFGLGAKSFNYIIRNFPNTVLIISHANT
  LSCDDYIVLRKGBEITVRCSEYEDVKTSQSELSHLLEMD"
  /complement(23896..31533)
  /gene="rRNA"
  /note="region containing small subunit, 5.8S and large
  subunit rRNA genes and spacer regions"
  /complement(23896..31533)
  /gene="rRNA"
  /complement(31966..32775)
  /gene="MAL1P3.04"
  /complement(31966..32775)
  /note="MAL1P3.04, conserved hypothetical membrane protein,
  len: 203 aa, similarity: P. falciparum chromosome 2,
  PF0010W, O96126 predicted integral membrane protein (255
  aa), fasta scores: opt: 335, E(): 4.9e-15, (36.1% identity
  in 191 aa overlap)"
  /codon_start=1
  /product="conserved hypothetical membrane protein,
  MAL1P3.04"
  /protein_id="CAB63559.1"
  /db_xref="GI:5594247"
  /translation="MKKSQYINVTILLFLLFLTYNYDTSKTKFNNNKKIDIN
  RFRKIIAASEEQYPMWEDFCLINEELPEHNDSPYLPHEHYENIDELNSINS
  TKIKWETIKKRONVYKEDTNHNNHNRDMHYKWKANYLYKVKHLINITLKDLTNPI
  HDKETITWIKIQEIDIEYFLNLOVEWLRILTELFYKKNKE"
  /complement(32477..32486)
  /gene="MAL1P3.04"
  /note="potential splice acceptor sequence"
  /complement(32669..32674)
  /gene="MAL1P3.04"
  /note="potential splice donor sequence, aaa/gtatat"
  /complement(36657..37343)
  /gene="MAL1P3.05"
  /complement(36657..37343)
  /gene="MAL1P3.05"
  /note="MAL1P3.05, hypothetical protein, len: 188 aa"
  /codon_start=1
  /product="hypothetical protein, MAL1P3.05"
  /protein_id="CAB63560.1"
  /db_xref="GI:5594248"
  /translation="MRKMNISGIFIKLLICISFCIFECFNKCMISYRKOLLWYSEN
  CFNYSIDRSLAEGSSSEKTKVKDIPNTELLKSLNINYEYKKNKEIVGSPMDNNLN

```


Qy 803 ataaaaatatctacacattgaccattgaaaaagtctcgttctcccatgggtaccagatcaa 862

Best Local Similarity 47.3%; Pred. No. 0.0019;
Matches 534; Conservative 0; Mismatches 574; Indels 20; Gaps


```
* 5492 6373: contig of 882 bp in length.
* 6374 6433: gap of unknown length
* 6434 7449: contig of 996 bp in length
* 7450 7530: gap of unknown length
* 7531 8140: contig of 610 bp in length
* 8141 8220: gap of unknown length
* 8221 9316: contig of 1096 bp in length
* 9317 10239: gap of unknown length
* 10240 10319: contig of 844 bp in length
* 10320 10964: gap of unknown length
* 10965 11044: contig of 645 bp in length
* 11045 11648: gap of unknown length
* 11649 11728: contig of 604 bp in length
* 11729 12690: gap of unknown length
* 12691 12776: contig of 968 bp in length
* 12777 13976: gap of unknown length
* 13977 14056: contig of 1200 bp in length
* 14057 15045: gap of unknown length
* 15046 15125: contig of 989 bp in length
* 15126 15969: gap of unknown length
* 15970 16049: contig of 844 bp in length
* 16050 16859: gap of unknown length
* 16860 16939: contig of 810 bp in length
* 16940 17663: gap of unknown length
* 17664 17743: contig of 723 bp in length
* 17744 18767: gap of unknown length
* 18768 18847: contig of 1025 bp in length
* 18848 18909: gap of unknown length
* 18910 19889: contig of 962 bp in length
* 19890 21046: gap of unknown length
* 21047 21126: contig of 1157 bp in length
* 21127 21820: gap of unknown length
* 21821 21908: contig of 700 bp in length
* 21909 23735: gap of unknown length
* 23736 23815: contig of 1829 bp in length
* 23816 25556: gap of unknown length
* 25557 26792: contig of 1741 bp in length
* 26793 26872: gap of unknown length
* 26873 28359: contig of 1156 bp in length
* 28360 28439: gap of unknown length
* 28440 29898: contig of 1487 bp in length
* 29899 31836: gap of unknown length
* 31837 31916: contig of 1459 bp in length
* 31917 33427: gap of unknown length
* 33428 34568: contig of 1431 bp in length
* 34569 35754: gap of unknown length
* 35755 35834: contig of 1141 bp in length
* 35835 37816: gap of unknown length
* 37817 39641: contig of 1106 bp in length
* 39642 39721: gap of unknown length
* 39722 41136: contig of 1746 bp in length
* 41137 42477: gap of unknown length
* 42478 42557: contig of 1414 bp in length
* 42558 44209: gap of unknown length
* 44210 44300: contig of 1262 bp in length
* 44301 45922: gap of unknown length
* 45923 46002: contig of 1672 bp in length
* 46003 47999: gap of unknown length
* 48000 49982: contig of 1613 bp in length
* 49983 50062: gap of unknown length
* 50063 51360: contig of 1903 bp in length
* 51361 53101: gap of unknown length
* 53102 53181: contig of 1298 bp in length
* 53182 54926: contig of 1661 bp in length
* 54927 54926: contig of 1745 bp in length
```

```
* 54927 55006: gap of unknown length
* 55007 56937: contig of 1931 bp in length
* 56938 57017: gap of unknown length
* 57018 57606: contig of 589 bp in length
* 57607 57687: gap of unknown length
* 57688 58332: contig of 946 bp in length
* 58333 58712: gap of unknown length
* 58713 60613: contig of 1901 bp in length
* 60614 60693: gap of unknown length
* 60694 62727: contig of 2034 bp in length
* 62728 62807: gap of unknown length
* 62808 65311: contig of 2504 bp in length
* 65312 65391: gap of unknown length
* 65392 66685: contig of 1294 bp in length
* 66686 66765: gap of unknown length
* 66766 68830: contig of 2065 bp in length
* 68831 68910: gap of unknown length
* 68911 71103: contig of 2193 bp in length
* 71104 71183: gap of unknown length
* 71184 72193: contig of 1010 bp in length
* 72194 72274: gap of unknown length
* 72275 74138: contig of 1865 bp in length
* 74139 74218: gap of unknown length
* 74219 76236: contig of 2018 bp in length
* 76237 76316: gap of unknown length
* 76317 77913: contig of 1597 bp in length
* 77914 77993: gap of unknown length
* 77994 80808: contig of 2815 bp in length
* 80809 80888: gap of unknown length
* 80889 82776: contig of 1888 bp in length
* 82777 82856: gap of unknown length
* 82857 85682: contig of 2826 bp in length
* 85683 85762: gap of unknown length
* 85763 89308: contig of 3546 bp in length
* 89309 89389: gap of unknown length
* 89390 94000: contig of 4611 bp in length
* 94001 109466: gap of unknown length
* 109467 109547: contig of 15387 bp in length
* 109548 110183: gap of unknown length
* 110184 110263: contig of 637 bp in length
* 110264 110842: gap of unknown length
* 110843 110922: contig of 579 bp in length
* 110923 111636: gap of unknown length
* 111637 111716: contig of 714 bp in length
* 111717 112030: gap of unknown length
* 112031 112110: contig of 314 bp in length
* 112111 112674: gap of unknown length
* 112675 112754: contig of 564 bp in length
* 112755 113301: gap of unknown length
* 113302 113381: contig of 547 bp in length
* 113382 113979: gap of unknown length
* 113980 114059: contig of 598 bp in length
* 114060 114698: gap of unknown length
* 114699 114778: contig of 639 bp in length
* 114779 115187: gap of unknown length
* 115188 115267: contig of 409 bp in length
* 115268 115938: gap of unknown length
* 115939 116018: contig of 671 bp in length
* 116019 116539: gap of unknown length
* 116540 116619: contig of 521 bp in length
* 116620 116619: gap of unknown length
```

Query Match 5.7%; Score 89; DB 55; Length 161891;
Best Local Similarity 41.2%; Pred. No. 0.0021;

Matches 432; Conservative 0; Mismatches 615; Indels 2; Gaps 2;

```
Oy 95 aatgcgacttaataaaattttacatttaataagaattttttttatcaataaatt 154
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 139941 AATGTATATATATATATATATATATATATATATATATATATATATA 139882
Oy 155 aatttattagttttattagaataatttaattagaataattttgaatcccgatttcctcc 214
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 139881 TATATAATATATATATGAATATTAGATTATAAATATGAATATATATA 139822
```


GenCore version 4.3
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 2, 2000, 00:25:15 : Search time 2111.72 Seconds
(without alignments)
1501.462 Million cell updates/sec

Title: US-09-464-528-16
Perfect score: 719
Sequence: 1 agatcaaacatcatcaaaa.....gttttgaagtataaagatg 719

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 5247842 seqs, 2204914090 residues

Total number of hits satisfying chosen parameters: 10495684

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:
1: em_est1:*
2: em_est2:*
3: em_est3:*
4: em_est4:*
5: em_est5:*
6: em_est6:*
7: em_est7:*
8: em_est8:*
9: em_est9:*
10: em_est10:*
11: em_est11:*
12: em_est12:*
13: em_est13:*
14: em_est14:*
15: em_est15:*
16: em_est16:*
17: em_est17:*
18: em_est18:*
19: em_est19:*
20: gb_est1:*
21: gb_est2:*
22: gb_est3:*
23: gb_est4:*
24: gb_est5:*
25: gb_est6:*
26: gb_est7:*
27: gb_est8:*
28: gb_est9:*
29: gb_est10:*
30: gb_est11:*
31: gb_est12:*
32: gb_est13:*
33: gb_est14:*
34: gb_est15:*
35: gb_est16:*
36: gb_est17:*
37: gb_est18:*
38: gb_est19:*
39: gb_est20:*
40: gb_est21:*
41: gb_est22:*
42: gb_est23:*
43: gb_est24:*

44: gb_est25:*
45: gb_est26:*
46: gb_est27:*
47: gb_est28:*
48: gb_est29:*
49: gb_est30:*
50: gb_est31:*
51: gb_est32:*
52: em_est20:*
53: em_est21:*
54: em_est22:*
55: em_est23:*
56: em_est24:*
57: em_est25:*
58: em_est26:*
59: gb_est33:*
60: gb_est34:*
61: gb_est35:*
62: gb_est36:*
63: gb_est37:*
64: gb_est38:*
65: em_est27:*
66: em_est28:*
67: em_est29:*
68: em_est30:*
69: gb_est39:*
70: gb_est40:*
71: gb_est41:*
72: gb_est42:*
73: gb_est43:*
74: gb_est44:*
75: em_est31:*
76: em_est32:*
77: em_est33:*
78: em_est34:*
79: gb_est45:*
80: gb_est46:*
81: gb_est47:*
82: em_est35:*
83: em_est36:*
84: em_est37:*
85: gb_est48:*
86: gb_est49:*
87: gb_est50:*
88: gb_est51:*
89: gb_est52:*
90: gb_est53:*
91: gb_est54:*
92: gb_est55:*
93: gb_gss1:*
94: gb_gss2:*
95: gb_gss3:*
96: gb_gss4:*
97: em_gss1:*
98: em_gss2:*
99: em_gss3:*
100: em_gss4:*
101: gb_gss5:*
102: gb_gss6:*
103: gb_gss7:*
104: gb_gss8:*
105: gb_gss9:*
106: em_gss5:*
107: em_gss6:*
108: em_gss7:*
109: em_gss8:*
110: em_gss9:*
111: em_gss10:*
112: em_gss11:*
113: gb_gss10:*
114: gb_gss11:*
115: em_gss12:*
116: gb_gss12:*

117: gb_gss13:.*
 118: gb_gss14:.*
 119: gb_gss15:.*
 120: gb_gss16:.*
 121: gb_gss17:.*
 122: gb_gss18:.*
 123: gb_gss19:.*
 124: em_gss13:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	83.6	11.6	843	122	CNS00CS1
C 2	82.8	11.5	1101	123	CNS016LI
C 3	81.2	11.3	1201	123	CNS016LI
C 4	81	11.3	1043	123	CNS0145P
C 5	79.8	11.1	1200	123	CNS016CO
C 6	78.8	11.0	928	122	CNS00DKY
C 7	77.6	10.8	1101	122	CNS00EVL
C 8	77.2	10.7	1101	122	CNS0021J
C 9	76.8	10.7	1101	122	CNS00FVG
C 10	76.8	10.7	1101	122	CNS00YVL
C 11	75.6	10.5	1101	122	CNS00EVL
C 12	74.6	10.4	928	122	CNS00DKY
C 13	74.6	10.4	1101	122	CNS001FB
C 14	74.4	10.3	996	122	CNS00FUG
C 15	74.4	10.3	1101	122	CNS003DQ
C 16	74.2	10.3	1043	123	CNS0145P
C 17	73.8	10.3	1225	123	CNS0161D
C 18	73.4	10.2	987	123	CNS014FQ
C 19	73.2	10.2	994	123	CNS015XG
C 20	72	10.2	1101	122	CNS003BD
C 21	72	10.0	1101	122	CNS00E07
C 22	71.6	10.0	1101	122	CNS0022H
C 23	71.4	9.9	770	113	AQ740708
C 24	71.4	9.9	828	113	AQ739398
C 25	71.4	9.9	987	123	CNS014FQ
C 26	71.2	9.9	820	120	B11728
C 27	71.2	9.9	867	122	CNS00CX5
C 28	71.2	9.9	1201	123	CNS0167M
C 29	70.6	9.8	1101	122	CNS00B01
C 30	70.2	9.8	905	122	CNS00KHX
C 31	70	9.7	843	122	CNS0091L
C 32	70	9.7	843	122	CNS00CS1
C 33	69.6	9.7	1101	122	CNS00LVZ
C 34	69.4	9.7	1101	122	CNS0042W
C 35	69	9.6	1101	122	CNS0022U
C 36	68.8	9.6	1101	122	CNS00D77
C 37	68.6	9.5	1187	120	B11102
C 38	68.6	9.5	1201	123	CNS01522
C 39	68.6	9.5	1225	123	CNS0161D
C 40	68.2	9.5	990	122	CNS00601
C 41	68.2	9.5	1101	122	CNS00Y2S
C 42	68	9.5	734	122	CNS010MP
C 43	68	9.5	1101	122	CNS000B8
C 44	68	9.5	1101	122	CNS00FMC
C 45	68	9.5	1101	122	CNS012BM

ALIGNMENTS

RESULT 1
 LOCUS CNS00CS1/c 843 bp DNA GSS 04-JUN-1999
 DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC # BAC26H19 of RPCI-98 library from Drosophila melanogaster (fruit

fly), genomic survey sequence.

AL059666
 AL059666.1 GI:4947129
 GSS
 KEYWORDS
 SOURCE fruit fly.

ORGANISM Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 843)

Genoscope.

Direct Submission

AUTHORS

TITLE

JOURNAL

COMMENT

Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)

- Web : www.genoscope.cns.fr

Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP).

The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information

please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osogawa and

Aaron Mammeter in Pieter de Jong's laboratory in the Department of

Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial

ECORI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's

p1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or

filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

Location/Qualifiers

1.843

source

/organism="Drosophila melanogaster"

/db_xref="taxon:7227"

/clone-lib="RPCI-98"

/clone="BAC26H19"

/note="end : tet3"

BASE COUNT 131 a 102 c 259 g 141 t 210 others

ORIGIN

Query Match 11.6%; Score 83.6; DB 122; Length 843;

Best Local Similarity 39.2%; Pred. No. 2.1e-05;

Matches 120; Conservative 64; Mismatches 122; Indels 0; Gaps 0;

QY 1 agatcaaacctcaccaaacatgatatctctccatcacatcactaactatttt 60

DB 828 AWATAAWAT 769

QY 61 tgggttaataataatcatttttttaagatatatttaagaaataaagaatttttt 120

DB 768 TATWTATTTTWTATATATATATATATATATATATATATATATATATATAT 709

QY 121 aaaaatgataaaatataattattcatgattttttcatatcattgattttgataaataa 180

DB 708 AWAT 649

QY 181 ttt 240

DB 648 TTTWTAT 589

QY 241 tttcaaaagcattcatcatttaataacatttaaaataatttaataacagtagaatt 300

DB 588 WATATAAAAWWATATATATATATATATATATATATATATATATATATAT 529

QY 301 tcttgt 306

DB 528 WATWAT 523

RESULT 2

CNS016LI/c

LOCUS CNS016LI 1101 bp DNA GSS 26-JUL-1999

DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC
BACN16D22 of DrosBAC library from Drosophila melanogaster (fruit
fly), genomic survey sequence.

ACCESSION AL106896
VERSION AL106896.1 GI:5624374
KEYWORDS GSS.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 1101)

REFERENCE
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage ;
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a
collaboration with the European Drosophila Genome Project (EDGP -
http://www.edgp.ebi.ac.uk - This Drosophila melanogaster BAC
library (Dros BAC) was made by Alain Billaud at CEPH (Centre
d'Etude du Polymorphisme Humain) with funding provided by a MRC
project grant. The DNA was prepared from embryos by Alain Bucheton
and Genevieve Payan. It has been constructed in the vector
pBeloBAC11.

FEATURES Location/Qualifiers
1..1101
 ~~/organism="Drosophila melanogaster"~~
 ~~/plasmid="pBeloBAC11"~~
 ~~/db_xref="taxon:7227"~~
 ~~/clone_lib="DrosBAC"~~
 ~~/clone="BACN16D22"~~
 ~~/note="end : T7"~~

BASE COUNT	203 a . 220 c 84 g 158 t 436 others
ORIGIN	

Query Match 11.5% ; Score 82.8 ; DB 123 ; Length 1101 ;
Best Local Similarity 33.6% ; Pred. No. 2.8e-05 ;
Matches 113 ; Conservative 90 ; Mismatches 133 ; Indels 0 ; Gaps 0 ;

```

QY      68 aaattaacattatcttttaagatatattaagaataaaaggatttttcaaaaaaatg 127
DB      1085 AWTYYTWAYWWWWWTCMTWANTAAAWMAAAMMAAAWMAAAWATTTATATTTHW 1026
QY      128 tacaaaattatattcatgatgttttcacacattgatttgataataaataatatttt 187
DB      1025 TAATHATNTWCATAAHNAWNTTATATWAATTAATTAHTATHTATYTWTWHWATWY 966
QY      188 tttaattctttaaaaaagtgtgcgaacaccttatagacagtcttgtctgtttacaac 247
DB      965 WTTYMTATWAAAAAATASTAAAYWHATAATWAAWATAATTTWTWTWATWTTWTA 906
QY      248 agcattcatcttatacataaaaaatttttaactaacagtagaattcttgtg 307
DB      905 WTAATATWTAWAAATAATWMAHHMMWTAAYAWWWMAWTAATWTAATWTAATWAAWATT 846
QY      308 agtgggtgtggagtagcgcaacctggtcgaacgagagagagagagccagaccagaag 367
DB      845 WTAATATATWAAAATATATATWTTAAAAAANAATWAAWAAATAATWAAAAAATAAHHM 786
QY      368 acaaaataaaagtatgcacaaacaaacaaatca 403
DB      785 TAAAAWATAAAVAAAAAANAANAANAANAANAHA 750

```

RESULT 3
CNS016EI/c CNS016EI 1201 bp DNA GSS 26-JUL-1999
LOCUS Drosophila melanogaster genome survey sequence SP6 end of BAC
DEFINITION BACN15C13 of DrosBAC library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL106627

REFERENCE
AUTHORS
TITLE
JOURNAL

AUTHORS
TITLE Genoscope.
JOURNAL Direct Submission
 Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
 BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
 - Web : www.genoscope.cns.fr)
COMMENT
 Determination of this BAC-end sequence was carried out as part of a
 collaboration with the European Drosophila Genome Project (EDGP) -
 http://www.edgp.ebi.ac.uk - This Drosophila melanogaster BAC
 library (Dros BAC) was made by Alain Billaud at CEPH (Centre
 d'Etude du Polymorphisme Humain) with funding provided by a MRC
 project grant. The DNA was prepared from embryos by Alain Bucheton
 and Genevieve Payan. It has been constructed in the vector
 pBelOBAC11.

FEATURES
 source
 1..1101
 /organism="Drosophila melanogaster"
 /plasmid="pBelOBAC11"
 /db_xref="taxon:7227"
 /clone_lib="DrosBAC"
 /clone="BACN01G13"
 /note="end : SP6"
 BASE COUNT 424 a 119 c 129 g 248 t 181 others
 ORIGIN

Query Match 10.7%; Score 76.8; DB 122; Length 1101;
 Best Local Similarity 39.5%; Pred. No. 0.00029;
 Matches 113; Conservative 55; Mismatches 118; Indels 0; Gaps 0;
 QY 6 aaactcacatccaaacataacatggatctctccatccatcatactaatatttgggt 65
 Db 1069 ATAAATWNTATWTAAGWNNWTTATWATTAATTAATATTTTWTWTTATATWTHW 1010
 QY 66 taaatattatcatatttttaagataataaataaagaatttttttaaaaaa 125
 Db 1009 TAAATWNTATWTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 950
 QY 126 tgcataaataatcatcatgatttttcatacatgatttgatgataaataatattt 185
 Db 949 WTWYTTTTTTTATWTTATWTTATWTTATWTTATWTTATWTTATWTTATWTTATW 890
 QY 186 ttttatttcttaaaaaatgctgaagacacttatagacatagctctgttctgtttac 245
 Db 889 TWTWTTTTTTTATWTTATTAATTTTWTWTTTATWTTATWTTATWTTATWTTATWTT 830
 QY 246 aaaagcattcatatttaatacattaaaaataatttaataactaaca 291
 Db 829 TWTWNTATRTATTTTWTWTTATWTTATWTTATWTTATWTTATWTTATWTTATWTT 784

RESULT 11
CNS00EVL 1101 bp DNA GSS 04-JUN-1999
LOCUS Drosophila melanogaster genome survey sequence T7 end of BAC:
DEFINITION BACR29B23 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL069706
VERSION AL069706.1 GI:4949849
KEYWORDS fruit fly.
SOURCE Drosophila melanogaster
ORGANISM Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1101)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
 BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
 - Web : www.genoscope.cns.fr)
COMMENT
 Determination of this BAC-end sequence was carried out as part of a
 collaboration with the Berkeley Drosophila Genome Project (BDGP).
 The BDGP is constructing a physical map of the Drosophila.

melanogaster genome using these BACs. For further information
 please see <http://www.fruitfly.org> The BDGP Drosophila
 melanogaster BAC library was prepared by Kazutoyo Osoegawa and
 Aaron Mamoser in Pieter de Jong's laboratory in the Department of
 Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
 NY. The library is named RPCI-98 and was constructed by partial
 EcoRI digestion of Drosophila DNA provided by the BDGP from the
 isogenic strain Y2; cn bw sp, the same strain used for the BDGP's
 p1 and EST libraries. A more detailed description of the library
 and how to order individual BAC clones, the entire library, or
 filters for hybridization from the BACPAC Resource Center can be
 found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES
 Location/Qualifiers
 1..1101
 /organism="Drosophila melanogaster"
 /db_xref="taxon:7227"
 /clone_lib="RPCI-98"
 /clone="BACR29B23"
 /note="end : T7"
 BASE COUNT 419 a 91 c 60 g 299 t 232 others
 ORIGIN

Query Match 10.5%; Score 75.6; DB 122; Length 1101;
 Best Local Similarity 36.0%; Pred. No. 0.00047;
 Matches 141; Conservative 71; Mismatches 180; Indels 0; Gaps 0;
 QY 14 atccaaacataacatggatctctccatccatcatactaatatttgggttaaatatt 73
 Db 462 WTWMMMMWMAATWTAAGWNNWTTATWATTAATTAATTAATTAATTAATTAATTAAT 521
 QY 74 aatcattatttttaagataataaataaagaatttttttaaaaaatgtataaa 133
 Db 522 TWTWTTWTTATWTTATWTTATWTTATWTTATWTTATWTTATWTTATWTTATWTT 581
 QY 134 attatatttcatgatttttcatcattgattgattgataaataatatttttttaatt 193
 Db 582 ATATTATWTTATWTTATWTTATWTTATWTTATWTTATWTTATWTTATWTTATWTT 641
 QY 194 ttccttaaaaaatgctgaagacacttatagacatagctctgtcttcacaaagcat 253
 Db 642 TTAATTATWTTATTAATTAATWTTATWTTATWTTATWTTATTAATTAATTAATTAAT 701
 QY 254 tcacatttaatacatttaaaaaatatttaatacactagatagctctctctgtgaggt 313
 Db 702 AAAAATWTAATWTTATWTTATWTTATWTTATWTTATWTTATWTTATWTTATWTT 761
 QY 314 gtgggagtaggcacccctgcatgaaacagagagagagagagagagagagagagagaaat 373
 Db 762 ATATATWTTWTTATWTTATWTTATWTTATWTTATWTTATWTTATWTTATWTTATWTT 821
 QY 374 aaaaagtatgcaacaaacaaatcaaatcaaa 405
 Db 822 AAATAWATAWTTATWTTATWTTATWTTATWTTATWTTATWTTATWTTATWTTATWTT 853

RESULT 12
CNS00DKY 928 bp DNA GSS 04-JUN-1999
LOCUS Drosophila melanogaster genome survey sequence T7 end of BAC #
DEFINITION BACR27A24 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL071865
VERSION AL071865.1 GI:4948170
KEYWORDS fruit fly.
SOURCE Drosophila melanogaster
ORGANISM Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 928)
AUTHORS Genoscope.
TITLE Direct Submission

JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)

COMMENT - Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's PI and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES Location/Qualifiers
source
1..928
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="RPCI-98"
/clone="BACR27A24"
/note="end : 77"

BASE COUNT 262 a 70 c 84 g 321 t 191 others

ORIGIN

Query Match 10.4%; Score 74.6; DB 122; Length 928;
Best Local Similarity 32.8%; Pred. No. 0.0007;
Matches 96; Conservative 91; Mismatches 103; Indels 3; Gaps 1;

QY 32 taTctctaccacatacactaattatttgggttaataataataatcattatttttaagat 91
Db 557 TWT 616
QY 92 attaatgaagaattaaagattttttaaataataataataataataataataataatgatt 151
Db 617 WAAATATWTAT 673
QY 152 tttcaccatttgatttgatataataataataataataataataataataataataatgctga 211
Db 674 AATAAAAAATTTTAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAW 733
QY 212 agacactttatgacatgctgtctctttacaaagcattcattcattcattcattcatt 271
Db 734 WAAATATWT 793
QY 272 aaaaattattactaacagtagaactctctgtgagtggtggtggtggtggtggtggtggt 324
Db 794 AAAAAAATAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAW 846

RESULT 13
CNS001FB/c 1101 bp DNA GSS 03-JUN-1999
Drosophila melanogaster genome survey sequence TET3 end of BAC #
BACR04A23 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.

ACCESSION AL060732
VERSION AL060732.1 GI:4939397
KEYWORDS GSS.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1101)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)

Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's PI and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES Location/Qualifiers
source
1..1101
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="RPCI-98"
/clone="BACR04A23"
/note="end : TET3"

BASE COUNT 288 a 110 c 103 g 491 t 109 others

ORIGIN

Query Match 10.4%; Score 74.6; DB 122; Length 1101;
Best Local Similarity 39.9%; Pred. No. 0.00069;
Matches 163; Conservative 55; Mismatches 184; Indels 7; Gaps 1;

QY 65 ttaaatattacattcatttttaagattatttaagatttaagatttaagatttttttaaaaaa 124
Db 1000 TWAATAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAW 941
QY 125 atgtataataattatattcattcattttttcattcatttttttttttttttttttttttt 184
Db 940 AAAAAAATAWATAWATAWATAWATAWATAWATAWATAWATAWATAWATAWATAWATAW 888
QY 185 ttttttaatt 244
Db 887 ATATWAAATATWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAW 828
QY 245 caaaacattcattcattcattcattcattcattcattcattcattcattcattcattcatt 304
Db 827 AAAAAATWTTTTTTTAAAT 768
QY 305 gtgagtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggt 364
Db 767 WWMWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAW 708
QY 365 aagacaaataaaagcatgcaacaaatacaatacaatacaatacaatacaatacaatacaata 424
Db 707 AATAWATAWATAWATAWATAWATAWATAWATAWATAWATAWATAWATAWATAWATAW 473
QY 425 ctcaattggtgtgacattcattcattcattcattcattcattcattcattcattcattcatt 473
Db 647 GGGGGGGGGGGGHHYHWTWHTWHTWHTWHTWHTWHTWHTWHTWHTWHTWHTWHTWHT 599

RESULT 14
CNS00FUFH/c 996 bp DNA GSS 03-JUN-1999
Drosophila melanogaster genome survey sequence TET3 end of BAC :
BACR31021 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.

ACCESSION AL071063
VERSION AL071063.1 GI:4951105
KEYWORDS GSS.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;

REFERENCE 1 (bases 1 to 996)
 AUTHORS Genoscope.
 TITLE Direct Submission
 JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)
 COMMENT - Web : www.genoscope.cns.fr
 Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammosier in Pletier de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain Y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES
 source
 Location/Qualifiers
 1..996
 /organism="Drosophila melanogaster"
 /db_xref="taxon:7227"
 /clone_lib="RPCI-98"
 /clone="BACR31021"
 /note="end : TET3"

BASE COUNT 383 a 164 c 81 g 171 t 197 others
 ORIGIN

Query Match 10.38; Score 74.4; DB 122; Length 996;
 Best Local Similarity 41.08; Pred. No. 0.00075;
 Matches 109; Conservative 50; Mismatches 107; Indels 0; Gaps 0;
 QY 23 taacatgagatccctaccatacactaatatttttgggttaaatattatcattat 82
 Db 863 TTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 804
 QY 83 tttaagaataataaagaattttttaaataaataaataaataaataaata 142
 Db 803 AAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 744
 QY 143 ttcagatgttttcacattgatttgatgaataaataatttttttcaattc 202
 Db 743 TTTTATTTTWTATTTTWTATTTTWTATTTTWTATTTTWTATTTTWTAT 684
 QY 203 aatgttgaagacattattagacattgtctgtctgttttcaaaagcattc 262
 Db 683 TTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 624
 QY 263 aatcattaaaaaattttaacta 288
 Db 623 WTTAKWATTTWKAATTTWGAATW 598

RESULT 15
 CNS003DQ/c 1101 bp DNA GSS 03-JUN-1999
 LOCUS Drosophila melanogaster genome survey sequence TET3 end of BAC #
 DEFINITION BACR08109 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
 ACCESSION AL064580
 VERSION AL064580.1 GI:4941932
 KEYWORDS GSS.
 SOURCE fruit fly.
 ORGANISM Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 REFERENCE 1 (bases 1 to 1101)

AUTHORS Genoscope.
 TITLE Direct Submission
 JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)
 COMMENT - Web : www.genoscope.cns.fr

Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammosier in Pletier de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain Y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES
 source
 Location/Qualifiers
 1..1101
 /organism="Drosophila melanogaster"
 /db_xref="taxon:7227"
 /clone_lib="RPCI-98"
 /clone="BACR08109"
 /note="end : TET3"

BASE COUNT 291 a 51 c 117 g 404 t 238 others
 ORIGIN

Query Match 10.38; Score 74.4; DB 122; Length 1101;
 Best Local Similarity 44.18; Pred. No. 0.00074;
 Matches 178; Conservative 36; Mismatches 187; Indels 3; Gaps 1;
 QY 19 aacataacatgagatccctaccatacactaatattttgggttaaatcaatca 78
 Db 535 MACWTAAATTTTAAATTTAAATTTAAATTTTAAATTTTAAATTTTAA 476
 QY 79 ttatttttaagatattaagaataaataaagatttttttaaaaaatgtataa 138
 Db 475 CCCWTTTTTATATTATTAATW--WAATTAATTAATAATAATAATAATA 419
 QY 139 attattcagatgttttcacattgatttgatgaataaataatttttttaatt 198
 Db 418 AAAAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 359
 QY 199 aaaaaattgtcgaagacactattagacattgtctgttttcaaaagcattc 258
 Db 358 WAATATATAATAATAATAATAATAATAATAATAATAATAATAATAATA 299
 QY 259 atttaatacataaaaaataatttaatacagtagaattctctgtgagtggtg 318
 Db 298 ATATTATTTTWTATTAATTAATTAATAATAATAATAATAATAATAATA 239
 QY 319 agtaggcaacctgacattgaacagagagaagagagagagagagagagag 378
 Db 238 AATATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 179
 QY 379 gtatgcaacaacaatacaaaatcaaaagggcgaagggctgggggtt 422
 Db 178 AA

Search completed: September 2, 2000, 00:25:22
 Job time: 4346 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 2, 2000, 02:00:43 : Search time 111.97 Seconds
(without alignments)
883.296 Million cell updates/sec

Title: US-09-464-528-16
Perfect score: 719
Sequence: 1 agatcaactcacatccaaa.....gttttgaagtataaagatg 719

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 243080 seqs, 68777915 residues

Total number of hits satisfying chosen parameters: 486160

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents, NA: *
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq: *
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq: *
3: /cgn2_6/ptodata/1/ina/5C_COMB.seq: *
4: /cgn2_6/ptodata/1/ina/5D_COMB.seq: *
5: /cgn2_6/ptodata/1/ina/6_COMB.seq: *
6: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq: *
7: /cgn2_6/ptodata/1/ina/backfiles1.seq: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	56	7.8	19124	4	US-08-487-826B-13
2	54.2	7.5	1186	4	US-08-731-722-5
3	52.6	7.3	665	4	US-08-883-795A-36
4	52	7.2	8920	3	US-08-446-855A-1
5	51.4	7.1	19124	4	US-08-487-826B-13
6	51.2	7.1	3933	4	US-08-731-722-3
7	51.2	7.1	3933	4	US-08-731-722-3
8	51	7.1	665	4	US-08-883-795A-36
9	50.2	7.0	473	2	US-08-764-100-16
10	50.2	7.0	4970	2	US-08-764-100-14
11	50.2	7.0	4970	2	US-08-764-100-20
12	50.2	7.0	8920	3	US-08-446-855A-1
13	49.8	6.9	4467	2	US-08-565-907A-1
14	49.8	6.9	4467	4	US-08-910-551B-1
15	49.8	6.9	4467	4	US-08-909-425A-1
16	49.4	6.9	1611	7	5213972-6
17	49.2	6.8	319	1	US-07-593-657-14
18	49	6.8	4098	4	US-08-605-106-4
19	48.2	6.7	6768	2	US-08-107-755A-1
20	48.2	6.7	8457	1	US-07-991-867B-1
21	48.2	6.7	8457	4	US-08-544-332-1
22	47.8	6.6	12124	1	US-08-181-271A-36
23	47.8	6.6	12124	1	US-08-449-315-36
24	47.8	6.6	12124	1	US-08-444-803-36
25	47.8	6.6	12124	1	US-08-449-043-36
26	47.8	6.6	12124	2	US-08-456-265A-36

27 47.8 6.6 12124 2 US-08-455-416-36 Sequence 36, Appl
28 47.8 6.6 12124 2 US-08-455-244-36 Sequence 36, Appl
29 47.8 6.6 12124 2 US-08-454-876-36 Sequence 36, Appl
30 47.8 6.6 12124 3 US-08-457-364-36 Sequence 36, Appl
31 47.8 6.6 12124 3 US-08-456-262-36 Sequence 36, Appl
32 47.8 6.6 12124 3 US-08-456-240-36 Sequence 36, Appl
33 47.8 6.6 12124 3 US-08-455-736-36 Sequence 36, Appl
34 47.8 6.6 12124 4 US-08-971-217-36 Sequence 36, Appl
35 47.6 6.6 10395 1 US-08-245-809-3 Sequence 5, Appl
36 47.6 6.6 10396 1 US-08-245-809-5 Sequence 2, Appl
37 47.6 6.6 10798 6 PCT-US92-01385-2 Sequence 2, Appl
38 47.6 6.6 10798 6 PCT-US92-01385-4 Sequence 4, Appl
39 47.6 6.6 10965 1 US-08-107-748-4 Sequence 4, Appl
40 47.6 6.6 10965 6 PCT-US92-01385-4 Sequence 13, Appl
41 47.4 6.6 642 2 US-08-764-100-13 Sequence 7, Appl
42 47.4 6.6 643 2 US-08-764-100-7 Sequence 2, Appl
43 47.4 6.6 2993 2 US-08-764-100-2 Sequence 10, Appl
44 47.4 6.6 2993 2 US-08-764-100-10 Sequence 9, Appl
45 47.4 6.6 3000 2 US-08-764-100-9

ALIGNMENTS

RESULT 1
US-08-487-826B-13
Sequence 13, Application US/08487826B
Patent No. 593827

GENERAL INFORMATION:

APPLICANT: Sim, Kim L.
APPLICANT: Chitnis, Chetan
APPLICANT: Miller, Louis H.
APPLICANT: Peterson, David S.
APPLICANT: Su, Xin-zhaun
APPLICANT: Wellens, Thomas E.

TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX

TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS

NUMBER OF SEQUENCES: 45

CORRESPONDENCE ADDRESS:

ADDRESSES: Knobe Martens Olson & Bear

STREET: 620 Newport Center Drive 16th Floor

CITY: Newport Beach

STATE: California

COUNTRY: US

ZIP: 92660

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/487,826B

FILING DATE: 10-SEP-1993

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Israelsen, Ned

REGISTRATION NUMBER: 29,655

REFERENCE/DOCKET NUMBER: NIH121.001CPI

TELECOMMUNICATION INFORMATION:

TELEPHONE: (619) 235-8550

TELEFAX: (619) 235-0176

INFORMATION FOR SEQ ID NO: 13:

SEQUENCE CHARACTERISTICS:

LENGTH: 19124 base pairs.

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: CDNA

HYPOTHETICAL: NO

ANTI-SENSE: NO

US-08-487-826B-13


```

Query Match      7.8%; Score 56; DB 4; Length 19124;
Best Local Similarity 47.7%; Pred. No. 0.0053;
Matches 164; Conservative 0; Mismatches 180; Indels 0; Gaps 0;

QY 54 attcatttgggttaaatatcaatcattattttaaagatatatttaagaataattaaagat 113
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1264 ATTATAATATGTAATTTATATAAATAATATTTTGTATACATACAAACACTTAAGAAAC 1323

QY 114 ttttataaaaatgataaaaatatattattcatgatttttttcatacatatttgatttgata 173
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1324 TATACAATCTGGTATCTAATAGTATATATATAATATCTTTTATTATTAATTTGTTCTCT 1383

QY 174 ataaaatatttttttttttcttaaaaagtgtgcgaagacacttatagacatagct 233
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1384 CTTTTTTTTTTTTTTTAAATAAATAAATAATTAATATATTTTTTTTTCATAATTATATGAT 1443

QY 234 tgtctgtttcaaaagcattcatttaacatcacatataaaaataatttaactaaacgt 293
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1444 TTAGTATTTTATAATAAATAAATCTTTTAAAAAACTTCAAAAACATTTTTCATAAATA 1503

QY 294 agaattcttcttgagtggtggagtaggcaacctggcattgaaacgagagagagaga 353
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1504 ATATTATATTAGTACCACCTAGTAAATTTAGAGAGAAACCTAGAACATACCCAAAAAA 1563

QY 354 gtcgaacacagagcaaaataaaaagtgtgcacaaacaaatca 397
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1564 ATTAGAACAAGAAAGATATTACAAAAAATAATAAAAAATTAAATTA 1607

RESULT 2
US-08-731-722-5/c
; Sequence 5, Application US/08731722
; Patent No. 5961971
; GENERAL INFORMATION:
; APPLICANT: Martin, Frank N.
; TITLE OF INVENTION: Biocontrol of Fungal Soilborne Pathogens
; TITLE OF INVENTION: by Pythium oligandrum
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Saliwanchik & Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: US
; ZIP: 32606-6669
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/731,722
; FILING DATE:
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Whitlock, Ted W.
; REGISTRATION NUMBER: 36,965
; REFERENCE/DOCKET NUMBER: UF-161
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 352-375-8100
; TELEFAX: 352-372-5800
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1186 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: 23-5
;
US-08-731-722-5

```


; Sequence 36, Application US/08883795A

Patent No. 5985607
GENERAL INFORMATION:
APPLICANT: Delcuve, Genevieve
TITLE OF INVENTION: Recombinant DNA Molecules and Expression
TITLE OF INVENTION: Vectors for Tissue Plasminogen Activator
NUMBER OF SEQUENCES: 39
CORRESPONDENCE ADDRESS:
ADDRESSEE: BERESKIN & PARR.
STREET: 40 King Street West
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5H 3Y2
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/883,795A
FILING DATE: 27-JUN-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Gravelle, Michelle
REGISTRATION NUMBER: 40,261
REFERENCE/DOCKET NUMBER: 7841-062
TELEPHONE: (416) 364-7311
TELEFAX: (416) 361-1398
INFORMATION FOR SEQ ID NO: 36:
SEQUENCE CHARACTERISTICS:
LENGTH: 665 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
IMMEDIATE SOURCE:
CLONE: Rh 32
US-08-883-795A-36

Query Match 7.1%; Score 51; DB 4; Length 665;
Best Local Similarity 53.0%; Pred. No. 0.095;
Matches 131; Conservative 0; Mismatches 115; Indels 1; Gaps 1;

Qy 46 tcaataatttttgggttaataatcaatttttttaagatatttaataaagaat 105
Db 508 TAAATTATAAATCTTTAATTATAAATATGTAATATAAATCTTTATAAATATGTAAT 449
Qy 106 taaagattttttaaataaataatgataaataattattcatgatttttcacacattga 165
Db 448 TATAAATATGTAATATAAATCTTTAATTATAAATATGTAATATAAATCTTTAAT 389
Qy 166 ttttgataataatatttttttcaatttttcaataatgttgcagacatttata 225
Db 388 TATAAATATGTAATATAAATCTTTAATTATAAATATGTAATATAAATCTTTAAT 329
Qy 226 catagctt-tgttcgtttacaagacattcatcatttaataacatttaataattt 284
Db 328 TATAAATATGTAATATAAATCTTTAATTATAAATATGTAATATAAATCTTTAAT 269
Qy 285 actaaca 291
Db 268 TATAAAT 262

RESULT 9
US-08-764-100-16
; Sequence 16, Application US/08764100
; Patent No. 5773700

GENERAL INFORMATION:
APPLICANT: van Grinsven J., Martinus Q.
APPLICANT: De Haan, Petrus T.
APPLICANT: Gielen L., Johannes J.
APPLICANT: Peters, Dirk
APPLICANT: Goldbach, Robert W.
TITLE OF INVENTION: Improvements in or Relating to Organic
TITLE OF INVENTION: Compounds
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sandoz Agro, Inc
STREET: 975 California Avenue
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/764,100
FILING DATE: 06-DEC-1996
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/214,064
FILING DATE:
APPLICATION NUMBER: US 08/032,235
FILING DATE: 17-MAR-1993
APPLICATION NUMBER: GB 9206016.9
FILING DATE: 19-MAR-1992
ATTORNEY/AGENT INFORMATION:
NAME: No. 5773700ris, Allen E.
REGISTRATION NUMBER: 34,490
REFERENCE/DOCKET NUMBER: 137-1061
TELEPHONE: (415) 354-3592
TELEFAX: (415) 857-1125
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 473 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-764-100-16

Query Match 7.0%; Score 50.2; DB 2; Length 473;
Best Local Similarity 52.1%; Pred. No. 0.047;
Matches 112; Conservative 0; Mismatches 103; Indels 0; Gaps 0;

Qy 67 aaataataatcatttttttaagatatatttaagaataataaagaatttttttaaaaaat 126
Db 236 AAAACCAAAAAATTTTGTGTAAATAAATAAGGCTCGGCCAGATTGGTCTAAGACC 295
Qy 127 gtataaaatattatttcattgatttttttcacacatttttgataataataattttt 186
Db 296 TTTTATTTGTTTTTATACATTTTATTGTTTGTGTTTTTATTATTATTATTATTTTA 355
Qy 187 ttttaattcttaaaaaatgttgcaagacacttattagacacattctgtctctgttaca 246
Db 356 TATTTTTTATATAGTTGCTTATTTAACACTTATTTAGACAAATTAATTTATTGATTA 415
Qy 247 aaagcattcatcatttaataacataaaaaatttt 281
Db 416 CAATCATCTGCTTATTTAATTTAAACACATTT 450

RESULT 10
US-08-764-100-14
; Sequence 14, Application US/08764100
; Patent No. 5773700

GENERAL INFORMATION:
APPLICANT: van Grinsven J., Martinus Q.
APPLICANT: De Haan, Petrus T.
APPLICANT: Gielen L., Johannes J.
APPLICANT: Peters, Dirk
APPLICANT: Goldbach, Robert W.
TITLE OF INVENTION: Improvements in or Relating to Organic
TITLE OF INVENTION: Compounds
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sandoz Agro, Inc
STREET: 975 California Avenue
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/764,100
FILING DATE: 06-DEC-1996
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/214,064
FILING DATE:
APPLICATION NUMBER: US 08/032,235
FILING DATE: 17-MAR-1993
APPLICATION NUMBER: GB 9206016.9
FILING DATE: 19-MAR-1992
ATTORNEY/AGENT INFORMATION:
NAME: No. 57737001s, Allen E.
REGISTRATION NUMBER: 34,490
REFERENCE/DOCKET NUMBER: 137-1061
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 354-3592
TELEFAX: (415) 857-1125
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 4970 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-764-100-14

Query Match 7.0%; Score 50.2; DB 2; Length 4970;
Best Local Similarity 52.1%; Pred. No. 0.06;
Matches 112; Conservative 0; Mismatches 103; Indels 0; Gaps 0;
QY 67 aaatattacatttttaagatatttaaaagataaaagatttttaaaaaaat 126
DB 1233 AAACCAAAAATTTTTTTTGTAAATAAAGGCTCGGCCAGATTGGCTAGACC 1292
QY 127 gataaaattattattacatttttttgcacacattatttagacagcttctgtttaca 246
DB 1293 TTTTATTTGTTTTATACATTTTATTGTTGTTGTTTATTATTATTATTTTA 1352
QY 187 ttttaattcttaaaaaatgttgcaagacacattatttagacagcttctgtttaca 246
DB 1353 TATTTTATATAGTTGCTTATTATTAACACTTATTATAGACAAATTAATTTTATTGATTA 1412
QY 247 aaagcattcatttaataacattaaaaaatattt 281
DB 1413 CAATCAATTCGCTTATTATTAATTTAAACACATTT 1447

RESULT 11
US-08-764-100-20/c
; Sequence 20, Application US/08764100
; Patent No. 5773700

GENERAL INFORMATION:
APPLICANT: van Grinsven J., Martinus Q.
APPLICANT: De Haan, Petrus T.
APPLICANT: Gielen L., Johannes J.
APPLICANT: Peters, Dirk
APPLICANT: Goldbach, Robert W.
TITLE OF INVENTION: Improvements in or Relating to Organic
TITLE OF INVENTION: Compounds
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sandoz Agro, Inc
STREET: 975 California Avenue
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/764,100
FILING DATE: 06-DEC-1996
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/214,064
FILING DATE:
APPLICATION NUMBER: US 08/032,235
FILING DATE: 17-MAR-1993
APPLICATION NUMBER: GB 9206016.9
FILING DATE: 19-MAR-1992
ATTORNEY/AGENT INFORMATION:
NAME: No. 57737001s, Allen E.
REGISTRATION NUMBER: 34,490
REFERENCE/DOCKET NUMBER: 137-1061
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 354-3592
TELEFAX: (415) 857-1125
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 4970 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-764-100-20

Query Match 7.0%; Score 50.2; DB 2; Length 4970;
Best Local Similarity 52.1%; Pred. No. 0.06;
Matches 112; Conservative 0; Mismatches 103; Indels 0; Gaps 0;
QY 67 aaatattacatttttaagatatttaaaagataaaagatttttaaaaaaat 126
DB 3738 AAACCAAAAATTTTTTTTGTAAATAAAGGCTCGGCCAGATTGGCTAGACC 3679
QY 127 gataaaattattattacatttttttgcacacattatttagacagcttctgtttaca 246
DB 3678 TTTTATTTGTTTTATACATTTTATTGTTGTTGTTTATTATTATTATTTTA 3619
QY 187 ttttaattcttaaaaaatgttgcaagacacattatttagacagcttctgtttaca 246
DB 3618 TATTTTATATAGTTGCTTATTATTAACACTTATTATAGACAAATTAATTTTATTGATTA 3559
QY 247 aaagcattcatttaataacattaaaaaatattt 281
DB 3558 CAATCAATTCGCTTATTATTAATTTAAACACATTT 3524

RESULT 12
US-08-446-855A-1
; Sequence 1, Application US/08446855A
; Patent No. 5849573

GENERAL INFORMATION:
APPLICANT: Stewart, Thomas S
APPLICANT: Flores, Maria V
APPLICANT: O'Sullivan, William J
TITLE OF INVENTION: Nucleotide sequence encoding carbamoyl
TITLE OF INVENTION: phosphate synthetase II
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Nixon & Vanderhye PC
STREET: 1100 NO. 5849573th Glebe Road, 8th Floor
CITY: Arlington
STATE: Virginia
COUNTRY: USA
ZIP: 22201-4714
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.24
CURRENT APPLICATION DATA: US/08/446.855A
APPLICATION NUMBER: US/08/446.855A
FILING DATE: 06-Jul-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Mitchard, Leonard C
REGISTRATION NUMBER: 29.009
REFERENCE/DOCKET NUMBER: 47-80
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-816-4000
TELEFAX: 703-816-4100
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 8920 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: genomic
US-08-446-855A-1

[illegible]

RESULT 13
US-08-565-907A-1/c
: Sequence 1, Application US/08565907A
: Patent No. 5814499
: GENERAL INFORMATION:
: APPLICANT: Sylvain Molneau, Barbara
: APPLICANT: J. Holler, Peter A. Vandenbergh,
: APPLICANT: Ebenezer R. Vedamuthu, Jeffrey K.

APPLICANT: Kondo
 TITLE OF INVENTION: DNA Encoding Phage
 TITLE OF INVENTION: Abortive Infection Protein
 TITLE OF INVENTION: From Lactococcus
 TITLE OF INVENTION: lacticis, and Method of Use Thereof
 NUMBER OF SEQUENCES: 1
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Ian C. McLeod
 STREET: 2190 Commons Parkway
 CITY: Okemos
 STATE: Michigan
 COUNTRY: USA
 ZIP: 48864
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette 5.25 inch,
 MEDIUM TYPE: 360 KB storage
 COMPUTER: Acer
 OPERATING SYSTEM: MS-DOS (version 4)
 SOFTWARE: Wordperfect 5.1
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/565,907A
 FILING DATE: December 1, 1995
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER:
 FILING DATE:
 ATTORNEY/AGENT INFORMATION:
 NAME: Ian C. McLeod
 REGISTRATION NUMBER: 20,931
 REFERENCE/DOCKET NUMBER: Quest 4.1-152
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (517) 347-4100
 TELEFAX: (517) 347-4103
 TELEX: NO. 5814499e
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 4467
 TYPE: Nucleotide
 STRANDEDNESS: Single
 TOPOLOGY: Linear
 MOLECULE TYPE:
 DESCRIPTION: Genomic DNA
 HYPOTHETICAL: No
 ANTI-SENSE: No
 FRAGMENT TYPE: N/A
 ORIGINAL SOURCE:
 ORGANISM: Lactococcus lactis
 STRAIN:
 INDIVIDUAL ISOLATE: W1
 DEVELOPMENTAL STAGE: N/A
 HAPLOTYPE: N/A
 TISSUE TYPE: N/A
 CELL TYPE: bacterium
 CELL LINE: N/A
 ORGANELLE: N/A
 IMMEDIATE SOURCE:
 LIBRARY: genomic
 CLONE: SMQ-20
 POSITION IN GENOME: N/A
 FEATURE:
 NAME/KEY: phage abortive infection
 LOCATION: N/A
 IDENTIFICATION METHOD: sequencing
 OTHER INFORMATION: DNA encoding phage
 OTHER INFORMATION: resistance
 PUBLIC INFORMATION: N/A
 IS-08-565-907A-1

```
Query Match      6.9%; Score 49.8; DB 2; Length 4467; *
Best Local Similarity 49.4%; Pred. NO. 0.071;
Matches 129; Conservative 0; Mismatches 132; Indels 0; Gaps 0;
```


[illegible]

GenCore version 4.5
Copyright (c) 1993 - 2000. Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 2, 2000, 02:07:29 ; Search time 161.21 Seconds
(without alignments)
1115.862 Million cell updates/sec

Title: US-09-464-528-16
Perfect score: 719
Sequence: 1 agatcaactcacatcaaa.....gttttgaagataaagatg 719

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 311585 seqs, 125096042 residues

Total number of hits satisfying chosen parameters: 623170

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_36.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	60.2	8.4	9789	1 T41852	cDNA encoding Plas
2	56.4	7.8	5849	1 V33135	Plasmodium berghei
3	56	7.8	4590	1 N60472	Sequence encoding
4	56	7.8	19124	1 T72882	Plasmodium var-7 g
5	54.2	7.5	1186	1 V29580	Pythium oligandrum
6	53.2	7.4	53585	1 X20251	Borrelia burgdorfe
7	52.6	7.3	605	1 T31530	Human 3' apolipop
8	52.4	7.3	26811	1 X20253	Borrelia burgdorfe
9	52.2	7.3	1826	1 V37413	Orpinomyces cellul
10	52.2	7.3	1826	1 V29477	Orpinomyces cellul
11	52.2	7.3	2503	1 Q33480	pNFX30 xylanase CD
12	52	7.2	8920	1 Q52924	Carbamoyl-phosphat
13	51.8	7.2	605	1 T31530	Human 3' apolipop
14	51.8	7.2	3975	1 N81157	Malaria-specific g
15	51.8	7.2	3975	1 Q22999	SERP gene. Recombi
16	51.8	7.2	6124	1 Q33568	Sequence encoding
17	51.8	7.2	26811	1 X20253	Borrelia burgdorfe
18	51.4	7.1	9789	1 T41852	cDNA encoding Plas
19	51.4	7.1	19124	1 T72882	Plasmodium var-7 g
20	51.2	7.1	783	1 X20361	Borrelia burgdorfe
21	51.2	7.1	1711	1 V33136	Plasmodium berghei
22	51.2	7.1	1864	1 N71405	Sequence of ANS-1
23	51.2	7.1	3933	1 V29578	Pythium oligandrum
24	51.2	7.1	3933	1 V29578	Pythium oligandrum
25	51	7.1	1470	1 Q55185	MS-Lei610 Vector.
26	50.8	7.1	110000	1 V21209_11	Continuation (12 o
27	50.2	7.0	2418	1 Q27886	P.falciplurum GBP13
28	50.2	7.0	4970	1 Q49959	Impatiens Necrotic
29	50.2	7.0	6124	1 Q03568	Sequence encoding
30	50.2	7.0	8920	1 Q62924	Carbamoyl-phosphat
31	50	7.0	2104	1 Q25273	Sequence encoding
32	49.8	6.9	4467	1 T68648	PSRQ800 fragment 1
33	49.4	6.9	1611	1 Q12528	Thymidylate phosph

34	49.4	6.9	53585	1 X20251	Borrelia burgdorfe
35	49.2	6.8	1240	1 Q24177	Tox2a gene. DNA en
36	49.2	6.8	2503	1 O53480	pNFX30 xylanase CD
37	49	6.8	1671	1 Q24134	50 kD subunit of S
38	49	6.8	4098	1 T43682	Medium chain-speci
39	48.8	6.8	6152	1 T78867	P. falciplurum live
40	48.8	6.8	110000	1 V21209_12	Continuation (13 o
41	48.6	6.8	6152	1 T78867	P. falciplurum live
42	48.6	6.8	110000	1 V21209_13	Continuation (14 o
43	48.4	6.7	1982	1 N90225	Malaria-specific p
44	48.4	6.7	2763	1 N95079	Gmshp26-A heat sho
45	48.2	6.7	2849	1 X20293	Borrelia burgdorfe

ALIGNMENTS

RESULT 1

T41852
ID T41852 standard; DNA: 9789 BP.
AC T41852;
DT 20-FEB-1997 (first entry)
DE cDNA encoding Plasmodium falciparum erythrocyte membrane protein.
KW Plasmodium falciparum; erythrocyte membrane protein; malaria;
KW detection; identification; treatment; prevention; parasite; ss.
OS Plasmodium falciparum MC type.
FH Key Location/Qualifiers
FT cds 326..9497
FT /*tag= a
FT /*product= Erythrocyte membrane protein
FT misc_feature 518..520
FT /*tag= b
FT /*transl_except= GTA encodes Tyrosine
FT misc_feature 656..658
FT /*tag= c
FT /*transl_except= ATT encodes Leucine
FT misc_feature 2909..2911
FT /*tag= d
FT /*transl_except= AAC encodes Aspartic acid
FT misc_feature 3461..3463
FT /*tag= e
FT misc_feature 5546..5548
FT /*tag= f
FT /*transl_except= CCT encodes Arginine
FT misc_feature 6254..6256
FT /*tag= g
FT misc_feature 6257..6259
FT /*tag= h
FT /*transl_except= ATA encodes Tyrosine
FT misc_feature 6263..6265
FT /*tag= i
FT misc_feature 6269..6271
FT /*tag= j
FT /*transl_except= TTC encodes Isoleucine
FT misc_feature 6272..6274
FT /*tag= k
FT misc_feature 6275..6277
FT /*tag= l
FT /*transl_except= ATT encodes Asparagine
FT misc_feature 6278..6280
FT /*tag= m
FT /*transl_except= GGA encodes Tryptophan
FT intron 7754..8478
FT /*tag= n

WO9633736-A1.

31-OCT-1996.

26-APR-1996; U05798.

27-APR-1995; US-430908.

PA (AFY-) AFFYMAX TECHNOLOGIES NV.

[illegible]

This is a partial nucleotide sequence of the mitochondrial DNA from a Pythium oligandrum isolate 23-5. Nucleotide sequences which are inverted repeats, flanked by PstI restriction sites from mitochondrial DNA from various P. oligandrum isolates are shown in V29576 to V29585. These Pythium isolate sequences are non-pathogenic and can be used as a method for controlling phytopathogenic organisms where the organisms are contacted with such a Pythium isolate. The Pythium isolates are used to control fungi, specifically pathogenic Pythium species, but possibly

CC other soil-borne pathogens), particularly for protecting plants
CC (seedlings, transplants or vegetable crops such as tomato) against
CC damping off, especially after transplanting into open fields. The
CC isolate sequences are useful as sources of probes for identification of
CC particular isolates. The isolates are widely distributed in nature, with
CC a similar ecology to pathogenic species. They produce large quantities of
CC oospore inoculum on liquid or solid substrates and are tolerant of
CC several commonly used fungicides. A single application at the greenhouse
CC stage will protect plants after transplanting.
SQ Sequence 1186 BP; 469 A; 121 C; 125 G; 471 T;
Query Match 7.5%; Score 54.2; DB 1; Length 1186;
Best Local Similarity 59.3%; Pred. NO. 0.38;
Matches 112; Conservative 0; Mismatches 73; Indels 4; Gaps 1;
QY 49 tactaatttttgggttaataatcattatttttaagataatttaataaagaattaa 108
DB 595 TATTGAAGAAATTTCTTATACTACTATTAAAGATTATAGGATCTACTGTATAAATCCAG 536
QY 109 aegatttttttaaaaaatgtataaaattattatttcattgatttttcatacatattgatt 168
DB 535 ATCATTTATTAAATATAAATAAAGTTT---TTCAATATTATAATAAATAATATATTC 480
QY 169 tgataataatatttttttttaattcttaaaaaatgtgcgaagacattatagacatt 228
DB 479 TTATTAGAGTATTTCATTTAAATTTTATTTTAAAGATTATATATCTTTTAAAAAGATAT 420
QY 229 agcttggtt 237
DB 419 AAATTAAT 411
RESULT 6
X20251/C
ID X20251 standard; DNA; 53585 BP.
AC X20251;
DE 04-MAY-1999 (first entry)
DE Borrelia burgdorferi polynucleotide sequence #4.
KW Borrelia burgdorferi; spirochete; bacterium; pathogen; Lyme disease;
KW epidemic relapsing fever; endemic relapsing fever; Lyme borreliosis;
KW infection; diagnosis; characterisation; detection; ds.
OS Borrelia burgdorferi.
PN W09858943-AL.
PD 30-DEC-1998.
PF 18-JUN-1998; U12764.
PR 03-SEP-1997; US-057483.
PR 20-JUN-1997; US-050359.
PR 22-JUL-1997; US-053344.
PR 22-JUL-1997; US-053377.
PA (HUMA-) HUMAN GENOME SCI INC.
PA (MEDI-) MEDIMUNE INC.
PI Clayton R, Dougherty BA, Fraser C, Lathigra R, Smith HO,
PI White OR;
DR WPI; 99-081217/07.
PT New isolated Borrelia burgdorferi nucleic acids - used to develop
PT products for the detection, diagnosis, characterisation, prevention
PT and therapy of infections, particularly Lyme disease
PS Claim 1; Page 801-831; 1128pp; English.
CC X20248 to X20402 represent polynucleotide sequences isolated from
CC Borrelia burgdorferi (Bb). Products derived from Bb can be used for
CC the detection, diagnosis, characterisation, prevention and therapy of
CC Bb infections, e.g. Lyme disease. They can also be used for the
CC production of biosynthetic products, e.g. enzymes. Borrelia belongs
CC to a family of motile, spiral-shaped bacteria called Spirochetes.
CC Spirochetes are pathogenic in humans and Borrelia causes epidemic and
CC endemic relapsing fever, and Lyme borreliosis, more commonly known as
CC Lyme disease.
SQ Sequence 53585 BP; 19235 A; 8081 C; 7001 G; 19266 T;
Query Match 7.4%; Score 53.2; DB 1; Length 53585;
Best Local Similarity 52.2%; Pred. NO. 0.46;

Matches 118; Conservative 0; Mismatches 108; Indels 0; Gaps 0;
QY 18 aacataacatggatgatactctcttaccatactactatttttgggttaaaatatttaac 77
DB 33384 AATATTAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATG 33325
QY 78 attatttttaagataataaataaagaataaagaatttttaaaaaatgtacaaatata 137
DB 33324 CAAATTTTGTAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATTT 33265
QY 138 tattattcattgatttttcatacatcattgattttgataataataatttttttaattct 197
DB 33264 TTTTAAAGTTTATTTTGGAAATAAATAAATAAATAAATAAATAAATAAATAAATA 33205
QY 198 taataaatgtgcgaagacattatttagacatagcttctgtctgttt 243
DB 33204 TTTTACCACCTAAGGAGCTCTATTATGAAAAACAGATTTTCTCTATAT 33159
RESULT 7
T31530
ID T31530 standard; cDNA; 605 BP.
AC T31530;
DT 15-SEP-1996 (first entry)
DE Human 3' apolipoprotein B SAR element clone Rh32.
KW Erythropoietin; EPO; anaemia; gene therapy; vector;
KW scaffold attachment region; SAR element; apolipoprotein B;
KW transgenic animal; ss.
OS Homo sapiens.
PN W09619573-AL.
PD 27-JUN-1996.
PF 18-DEC-1995; CA0696.
PR 19-DEC-1994; US-358918.
PA (CANG-) CANGENE CORP.
PI Delcuve G;
DR WPI; 96-309587/31.
PT Recombinant DNA molecule expressing mammalian erythropoietin
PT useful to transform cell lines, and for gene therapy, e.g. of
PT anaemia and other red blood cell disorders
PS Claim 7; Page 59-60; 84pp; English.
CC Human apolipoprotein B (apoB) scaffold attachment region (SAR)
CC element clones Rh32 (T31530) and Rh10 (T31531) respectively carry
CC the 3' human apoB SAR element and the distal 1212 bp 5' human apoB
CC SAR element and 1317 bp proximal sequence. These SAR elements
CC co-map with the boundaries of the human apoB gene chromatin
CC domain. A novel recombinant DNA molecule adapted for transfection
CC of a host cell comprises an erythropoietin (EPO) cDNA (T31529) or
CC genomic clone (T31532) operably linked to an expression control
CC sequence and to the 5' and 3' SAR elements. The SAR elements
CC increase expression of the recombinant EPO in stable, long-term
CC mammalian cell cultures.
SQ Sequence 605 BP; 278 A; 18 C; 35 G; 274 T;
Query Match 7.3%; Score 52.6; DB 1; Length 605;
Best Local Similarity 55.1%; Pred. NO. 0.7;
Matches 124; Conservative 0; Mismatches 99; Indels 2; Gaps 1;
QY 70 tattaatcattttttaagataataaagaataaagaatttttaaaaaatgta 129
DB 25 TATAATTAATAATTTATATAATTAATAATTTATTAATAAATAAATAAATAAATAATTT 84
QY 130 taataattattattcatgatttttcatacatcattgattttgataataataattttttt 189
DB 85 TATAATTAATAATTTATATAATTAATAAATAAATAAATAAATAAATAAATAAATAAT 142
QY 190 taatttttaaaaaatgtgcgaagacattatttagacatagcttctgtctgttcaaaaa 249
DB 143 TTTTATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAT 202
QY 250 gcattcatttaataacataaaaaataattttaactaacagta 294
DB 203 TTTTATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAT 247

```

RESULT      8
X20253
ID X20253 standard; DNA; 26811 BP.
AC X20253;
DT 04-MAY-1999 (first entry)
DE Borrelia burgdorferi polynucleotide sequence #6.
KW Borrelia burgdorferi; spirochete; bacterium; Lyme disease;
KW epidemic relapsing fever; endemic relapsing fever; Lyme borreliosis;
KW infection; diagnosis; characterisation; detection; ds.
OS Borrelia burgdorferi.
PN W09858943-A1.
PD 30-DEC-1998.
PF 18-JUN-1998: U12764.
PR 03-SEP-1997: US-057483.
PR 20-JUN-1997: US-050359.
PR 22-JUL-1997: US-053344.
PR 22-JUL-1997: US-053377.
PA (HUMA-) HUMAN GENOME SCI INC.
PA (MED-) IMMUNE INC.
PI Clayton R, Dougherty BA, Fraser C, Lathigra R, Smith HO,
PI White OR;
DR WPI: 99-081217/07.
PT New isolated Borrelia burgdorferi nucleic acids - used to develop
PT products for the detection, diagnosis, characterisation, prevention
PT and therapy of infections, particularly Lyme disease
PS Claim 1; Page 851-867; 1129pp; English.
CC X20248 to X20402 represent polynucleotide sequences isolated from
CC Borrelia burgdorferi (BB). Products derived from BB can be used for
CC the detection, diagnosis, characterisation, prevention and therapy of
CC Bb infections, e.g. Lyme disease. They can also be used for the
CC production of biosynthetic products, e.g. enzymes. Borrelia belongs
CC to a family of motile, spiral-shaped bacteria called Spirochetes.
CC Spirochetes are pathogenic in humans and Borrelia causes epidemic and
CC endemic relapsing fever, and Lyme borreliosis, more commonly known as
CC Lyme disease.
SQ Sequence 26811 BP; 8978 A; 3456 C; 3596 G; 10780 T;

Query Match      7.3%; Score 52.4; DB 1; Length 26811;
Best Local Similarity 52.3%; Pred. No. 0.63;
Matches 116; Conservative 0; Mismatches 106; Indels 0; Gaps 0;

QY 76 tcattatttttaagatatttaataaagaatttaaaagatttttttaaaatgtataaat 135
DB 14768 TAATTATTATTCTATTGTTAAATAGATTATTATTAGATTATTATTATTATTATAAATAA 14827

QY 136 tatattattcatgattttcacacattgtttgttgataataatatttttttaattt 195
DB 14828 TAATTATTATTGAATTTTAAATAATAGATTATTATTATTATTATTATTATTATTATTAA 14887

QY 196 cttaaaaaaagtgcgaagacattattagacatagctgtgtgtgttttcaaaagcattc 255
DB 14888 CACTTCCTTCTGGCAATAAAGTCTTATACAGAAGTCTTCTGCTTTTAAACACCTTT 14947

QY 256 atcatttaatacattaaaaatttttaataactaactaagtagaa 297
DB 14948 TAAAGAGTTTAAAGAACTTTCTTATATATATTCTTCAACAAAAGCA 14989

RESULT      9
V37413
ID V37413 standard; cDNA to mRNA; 1826 BP.
AC V37413;
DT 14-SEP-1998 (first entry)
DE Orpinomyces cellulase CelB cDNA.
KW Cellulase; endoglucanase; CelB gene; ds.
OS Orpinomyces sp. strain PC-2.
FH Key Location/Qualifiers
FT CDS
FT 69..1484
FT /*tag= a
FT /note= "Claim 5"
PT New recombinant DNA encoding Orpinomyces cellulase protein - useful

```

```

PN W09814597-A1.
PD 09-APR-1998.
PF 03-OCT-1997: U18008.
PR 04-OCT-1996; US-027883.
PA (UYGE-) UNIV GEORGIA RES FOUND INC.
PI Chen H, Li X, Ljungdahl LG;
DR WPI: 98-240096/21.
DR P-PSDB: W56742.
PT New recombinant DNA encoding Orpinomyces cellulase protein - useful
PT for, e.g. producing recombinant Orpinomyces cellulase in host cell
PS Claim 5; Page 38-40; 69pp; English.
CC This cDNA clone includes a claimed coding region for cellulase celB
CC (see W56742) of the anaerobic bovine rumen fungus Orpinomyces sp.
CC PC-2. It was obtained by screening a PC-2 cDNA library for clones
CC active on remazol brilliant blue-carboxymethylcellulose. The
CC encoded cellulase has endoglucanase, but not cellobiohydrolase,
CC activity. CelA and celC genes (see V29472-73), also obtained from
CC Orpinomyces sp. PC-2, encode cellulases having both activities
CC (see W56738-39). Recombinant DNA molecules encoding Orpinomyces
CC cellulase proteins are claimed, as well as recombinant cells
CC selected from Saccharomyces cerevisiae, Escherichia coli,
CC Aspergillus, Trichoderma reesei, Pichia, Penicillium, Streptomyces
CC or Bacillus, and a method for producing recombinant cellulase by
CC culturing these host cells.
SQ Sequence 1826 BP; 654 A; 236 C; 313 G; 623 T;

Query Match      7.3%; Score 52.2; DB 1; Length 1826;
Best Local Similarity 49.8%; Pred. No. 0.76;
Matches 132; Conservative 0; Mismatches 133; Indels 0; Gaps 0;

QY 10 tcacatccaaacaataacatggatctctaccatcatcatactattttgggttaa 69
DB 1501 TTAACAACATAAATAATTATTATTAGTAAATAAAGAAATAAATTTTAAATAATAT 1560

QY 70 tattaatcattttttaagatatttaataaagaatttaaaagatttttttaaaagatgta 129
DB 1561 ATTATATTATTGTTTATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 1620

QY 130 taaaattattattattcatgattttttcatcacattgtttgtgatacaataatattttt 189
DB 1621 AAATATATACATAACAAAAGTAAAAAATTTTAGTATTAGTATTAGTATTAGTATTAGT 1680

QY 190 taattttctaaaaaatgttcgaagacattattagacatagctgttctgtttacaaaa 249
DB 1681 AAAAAGTTTAATAAATAGTAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATA 1740

QY 250 gcattcatcatatttaacacattaa 274
DB 1741 AGAAATAAATAAATAAATACITTTAAA 1765

RESULT 10
V29477
ID V29477 standard; cDNA to mRNA; 1826 BP.
AC V29477;
DT 21-SEP-1998 (first entry)
DE Orpinomyces cellulase CelB cDNA.
KW Cellulase; endoglucanase; CelB gene; ds.
OS Orpinomyces sp. strain PC-2.
FH Key Location/Qualifiers
FT CDS
FT 69..1484
FT /*tag= a
FT /note= "Claim 5"
PT New recombinant DNA encoding Orpinomyces cellulase protein - useful

```

PT for, e.g. producing recombinant *Orpinomyces* cellulase in host cell

PS Claim 5; Page 38-40; 69pp; English.

CC This cDNA clone includes a claimed coding region for cellulase celB
CC (see W56742) of the anaerobic bovine rumen fungus *Orpinomyces* sp.
CC PC-2. It was obtained by screening a PC-2 cDNA library for clones
CC active on remazol brilliant blue-carboxymethylcellulose. The
CC encoded cellulase has endoglucanase, but not cellobiohydrolase,
CC activity. CelA and celC genes (see V29472-73), also obtained from
CC *Orpinomyces* sp. PC-2, encode cellulases having both activities
CC (see W56738-39). Recombinant DNA molecules encoding *Orpinomyces*
CC cellulase proteins are claimed, as well as recombinant cells
CC selected from *Saccharomyces cerevisiae*, *Escherichia coli*,
CC *Aspergillus*, *Trichoderma reesei*, *Pichia*, *Penicillium*, *Streptomyces*
CC or *Bacillus*, and a method for producing recombinant cellulase by
CC culturing these host cells.

SQ Sequence 1826 BP; 654 A; 236 C; 313 G; 623 T;

Query Match 7.3%; Score 52.2; DB 1; Length 1826;

Best Local Similarity 49.8%; Pred. No. 0.76;

Matches 132; Conservative 0; Mismatches 133; Indels 0; Gaps 0;

QY 10 tccatcccaaacataacatgagatctcctaccacatcactaactattttgggttaa 69

DB 1501 TTACAACATAATAAATTTAGTAAATAAAGAAATAAATTTTAAATAATAT 1560

QY 70 tattaatcatttttaagatttaagaaattaaagatttttttaaaaaaatgta 129

DB 1561 ATTATATATATGTTATTAATAATAATAATAATAATAATAATAATAATA 1620

QY 130 taaatattattatcatgatgtttttcaccatgtttgatttgataataatatttttt 189

DB 1621 AATATATATACATAACAAAGTAAAAATTTAAATTTTGTAGTATTGTATAAATTTATT 1680

QY 190 taattttcttaaaagtgtcagacacactcattagacatgcttctgttttcaaaaa 249

DB 1681 AAAAAGTTTAATAAGTAAATAAATAAATAAATAAATAAATAAATAAATAA 1740

QY 250 gcattcatcatttaacacatacaaa 274

DB 1741 AGAATAATAATAATACTTTTAAAA 1765

RESULT 11

Q53480

ID Q53480 standard; cDNA; 2503 BP.

AC Q53480; 30-JUN-1994 (first entry)

DE pNPX30 xylanase cDNA.

KW xylanase; ruminant animals; fungus; paper; pulp; bagasse;

KW feedstock; rumen; plant fibre; ss.

OS Neocallimastix patriciarum.

FH key Location/Qualifiers

FT cds 1.1935

FT /tag= a

FT /product= Xylanase.

PN W09325671-A.

PD 23-DEC-1993.

PF 17-JUN-1993; AU0294.

PR 17-JUN-1992; AU-002985.

PR 29-JUN-1992; AU-003238.

PR 01-APR-1993; AU-008100.

PA (CSIR) COMMONWEALTH SCI & IND RES ORG.

PI Xue G;

DR WPI; 94-007529/01.

DR P-PSDB; R44529.

PT New recombinant fungal xylanase - used for hydrolysis of xylan in

PT food and pulp and paper industries and for improving ruminant

PT feed efficiency

PS Claim 9; Figure 3; 45pp; English.

CC The cloned xylanase coding sequence is derived from an anaerobic

CC rumen fungus. The xylanase has high specific activity for the

CC hydrolysis of xylan. It can be used for treating pulps in the pulp

CC and paper industry, for treating bagasse for more efficient disposal
CC or for the treatment of feedstock to improve nutritional value.
CC Genetically modified xylanase genes can also be used for the
CC modification of rumen bacteria to improve plant fibre utilisation by
CC ruminants. 2503 BP; 893 A; 389 C; 517 G; 704 T;

SQ Sequence 2503 BP; 893 A; 389 C; 517 G; 704 T;

Query Match 7.3%; Score 52.2; DB 1; Length 2503;

Best Local Similarity 51.6%; Pred. No. 0.75;

Matches 144; Conservative 0; Mismatches 133; Indels 2; Gaps 1;

QY 18 aaacataacatgatctctcctaccacatcactaactattttgggttaaatattc 77

DB 2070 AAATTTTAAATGTAATAATTTTAAATAATACAAATTTGTAATAATAATGATG 2129

QY 78 attatttttaagattatttaagaaattaaagatttttttaaaaaatgtataaatta 137

DB 2130 AAAAATTTAAATGTAATAATTTTAAATAATACAAATTTGTAATAATAATGATG 2189

QY 138 tattattcatgatgtttttcaccatgtttgattttgataa--taaatatttttttaatt 195

DB 2190 AAAAATAAAGAATTTATGAATAATTTTAAATGTAATAATTTTAAATAATATATAT 2249

QY 196 cttaaaaaagtgtcagacacactcattagacatgcttctgttttcaaaaagcattc 255

DB 2250 AATAAAGAATTTATGAATAATTTTAAATAATATAAATGTTGTTTAAATAATA 2309

QY 256 atcatttaacatttaaaaaatttaataactaacagta 294

DB 2310 AAAAATTTATGAATAATTTTAAATAATAAATAAATAAATAAATA 2348

RESULT 12

Q62924/c standard; cDNA; 8920 BP.

ID Q62924; 06-DEC-1994 (first entry)

AC Q62924; 06-DEC-1994 (first entry)

DE Carbamoyl-phosphate-synthetase II.

KW Carbamoyl-phosphate-synthetase II; CPSII; pscPSII gene;

KW malaria; ss.

OS Plasmodium falciparum.

FH key Location/Qualifiers

FT cds 1226.8401

FT /tag= a

FT /EC_number= 6.3.5.5

PN W09412643-A.

PD 09-JUN-1994.

PF 02-DEC-1993; AU0617.

PR 03-DEC-1992; AU-006206.

PR 16-DEC-1992; AU-006380.

PA (UNIX) UNISEARCH LTD.

PI Flores MV, Osullivan WJ, Stewart TS;

DR WPI; 94-200271/24.

DR P-PSDB; R55694.

PT Nucleic acid encoding carbamoyl phosphate synthetase II -

PT isolated from Plasmodium falciparum, used to develop prods. for

PT the treatment of malaria.

PS Disclosure; Page 6-16; 31pp; English.

CC The cDNA sequence encoding the carbamoyl-phosphate-transferase II

CC (CPSII) of Plasmodium falciparum was determined. The cDNA encodes

CC a protein that includes 2 insert sequences not found in other CPSII

CC proteins. The first separates the putative structural subdomain and

CC the glutaminase subdomain of the glutamine-amidotransferase subunit

CC of CPSII, while the second separates 2 ATP binding subdomains of the

CC CPSII subunit, CPsA and CPsB.

SQ Sequence 8920 BP; 8936 A; 774 C; 1232 G; 3078 T;

Query Match 7.2%; Score 52; DB 1; Length 8920;

Best Local Similarity 49.8%; Pred. No. 0.76;

Matches 133; Conservative 0; Mismatches 135; Indels 0; Gaps 0;

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 2, 2000, 02:55:46 : Search time 5217.98 Seconds
(without alignments)
245.918 Million cell updates/sec

Title: US-09-464-528-16
Perfect score: 719

Sequence: 1 agatcaaacacacatccaaa.....gttttgaagtataaagatg 719

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 972840 seqs, 892348106 residues

Total number of hits satisfying chosen parameters: 1945680

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl:*

1: gb_ba1:*
2: gb_ba2:*
3: gb_cm:*
4: gb_ov:*
5: gb_pat:*
6: gb_ph:*
7: gb_pl1:*
8: gb_pl2:*
9: gb_pr1:*
10: gb_pr2:*
11: gb_pr3:*
12: gb_ro:*
13: gb_sts:*
14: gb_sy:*
15: gb_un:*
16: em_fun:*
17: em_hum1:*
18: em_hum2:*
19: em_in:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_sy:*
29: em_un:*
30: em_v1:*
31: gb_htg1:*
32: gb_htg2:*
33: gb_in1:*
34: gb_in2:*
35: em_ba1:*
36: em_ba2:*
37: em_hum3:*
38: em_hum4:*
39: gb_pr4:*
40: gb_htg3:*
41: gb_htg4:*
42: gb_htg5:*
43: gb_htg6:*

44: gb_htg7:*
45: em_htg1:*
46: em_htg2:*
47: em_htg3:*
48: em_hum5:*
49: gb_pl3:*
50: gb_br5:*
51: gb_htg8:*
52: gb_htg9:*
53: gb_htg10:*
54: gb_htg11:*
55: gb_htg12:*
56: gb_htg13:*
57: gb_htg14:*
58: gb_in3:*
59: gb_htg15:*
60: gb_htg16:*
61: gb_htg17:*
62: em_htg4:*
63: em_htg5:*
64: em_htg6:*
65: em_htg7:*
66: em_hum6:*
67: gb_htg18:*
68: gb_htg19:*
69: gb_htg20:*
70: gb_htg21:*
71: gb_htg22:*
72: gb_htg23:*
73: gb_htg24:*
74: gb_htg25:*
75: gb_htg26:*
76: gb_htg27:*
77: gb_htg28:*
78: gb_htg29:*
79: gb_htg30:*
80: gb_htg31:*
81: gb_v11:*
82: gb_v12:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	74	10.3	144759	10 HS352A20	AL021939 Homo sapi
2	70.2	9.8	153267	57 AC011212	AC011212 Homo sapi
3	70	9.7	104992	41 AC005504	AC005504 Plasmodiu
4	70	9.7	130281	60 AC004157	AC004157 Plasmodiu
5	69.8	9.7	349919	54 AC008576	AC008576 Homo sapi
6	69.6	9.7	164119	75 AC026640	AC026640 Homo sapi
7	69.2	9.6	179510	71 AC013820	AC013820 Homo sapi
8	69	9.6	48532	51 AC023371	AC023371 Homo sapi
9	68.4	9.5	975	33 DD087514	U87514 Dictyosteli
10	68.2	9.5	80920	60 AC006278	AC006278 Plasmodiu
11	68.2	9.5	121024	8 CHMPXX	X04465 Marchantia
12	67.8	9.4	910	13 CNS01G8P	AL142826 Anopheles
13	66.8	9.3	99263	32 CNS01DX9	AL139177 Homo sapi
14	66.8	9.3	152209	11 HS1108D11	AL034419 Human DNA
15	66.6	9.3	1867	7 MTSCA123	AJ223323 Saccharom
16	66.4	9.2	149752	11 AC004616	AC004616 Homo sapi
17	66.4	9.2	170427	41 AC006095	AC006095 Homo sapi
18	66.2	9.2	167271	53 AC006269	AC006269 Homo sapi
19	66.2	9.2	226345	39 AC005406	AC005406 Homo sapi
20	66	9.2	158131	68 AC009277	AC009277 Homo sapi
21	65.8	9.2	2982	33 DDISGSPA	M3862 Dictyosteli
22	65.8	9.2	67970	33 PFMALIP3	AL031746 Plasmodiu
23	65.4	9.1	690	7 MISC10	V00691 Two yeast m
24	65.4	9.1	716	8 YSCMTTGT1	J01533 Yeast (S.ce

25 65.2 9.1 68879 72 AC031982 Homo sapi
 c 26 65.2 9.1 287731 59 AC008620 Homo sapi
 27 65 9.0 2305 33 DMTRNA
 c 28 65 9.0 183638 77 AC009653 Homo sapi
 29 65 9.0 216406 52 AC018919 Homo sapi
 c 30 64.8 9.0 161891 55 AC008206 Drosophila
 c 31 64.8 9.0 163878 52 AC010178 Homo sapi
 c 32 64.8 9.0 173893 57 AC021553 Homo sapi
 c 33 64.4 9.0 217242 73 AC016222 Homo sapi
 34 64.2 8.9 153267 57 AC011212 Homo sapi
 c 35 64.2 8.9 174434 42 AC011115 Homo sapi
 c 36 64 8.9 1398 8 SCU46121 Saccharomyc
 c 37 64 8.9 2606 7 YSCMTG07 L38891 Saccharomyc
 38 64 8.9 14867 34 AE001398 Plasmodi
 39 64 8.9 85779 8 SC011856 Saccharom
 40 64 8.9 106650 39 AC007708 Homo sapi
 c 41 63.8 8.9 6591 7 YSCMTG06 L36890 Saccharomyc
 c 42 63.8 8.9 181650 79 AC009760 Homo sapi
 c 43 63.8 8.9 193372 53 AC022266 Homo sapi
 c 44 63.8 8.9 204951 60 AC005505 Plasmodi
 45 63.6 8.8 145932 59 AC024387 Homo sapi

ALIGNMENTS

RESULT 1

HS352A20 144759 bp DNA 23-NOV-1999
 LOCUS Homo sapiens DNA sequence from PAC 352A20 on chromosome
 DEFINITION 6q24.1-25.1. Contains a pseudogene similar to yeast, bacterial,
 worm and slime mold hypothetical genes, and a gene coding for an
 aldehyde dehydrogenase family protein. Contains ESTs, STSs and
 GSSs, complete sequence.

ACCESSION AL021939
 VERSION AL021939.1 GI:3135969
 KEYWORDS HTG: aldehyde dehydrogenase.
 SOURCE human.

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Euthera; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 144759)

PHILLIPS,S.
 Direct Submission
 Submitted (07-MAY-1998) sanger.ac.uk/HGP/Chr6/ Sanger Centre,
 Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquires:
 humquerry@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk

On May 18, 1998 this sequence version replaced gi:2909620.
 IMPORTANT: This sequence is the entire insert of clone 352A20.

During sequence assembly data is compared from overlapping clones.
 Where differences are found these are annotated as variations
 together with a note of the overlapping clone name. Note that the
 variations annotated may not be found in the sequence submission
 corresponding to the overlapping clone as we submit sequences with
 only a small overlap as described above.

This sequence was generated from part of bacterial clone contigs of
 human chromosome 6, constructed by the Sanger Centre chromosome 6
 mapping group. Further information can be found at
 http://www.sanger.ac.uk/HGP/Chr6/

This sequence has been finished according to sequence map criteria
 as follows. An attempt is made to resolve all sequencing problems,
 such as compressions and repeats, but not necessarily within known
 annotated human repeat sequence elements (e.g. Alu). Where the
 sequence is ambiguous, there is an annotation using the 'unsure'
 feature key.

The true left end of clone 352A20 is at 1 in this sequence. The
 true right end of clone 352A20 is at 144759.
 352A20 is from the library RPC13 constructed at the Roswell Park
 Cancer Institute by the group of Pieter de Jong.
 For further details see http://bacpac.med.buffalo.edu/.

FEATURES
 Source

1..144759
 /organism="Homo sapiens"

/db_xref="taxon:9606"
 /chromosome="6"
 /map="q24.1-25.1"
 /clone="RP3-352A20"
 /clone_lib="RPC1-3"
 3..211
 /note="AluSc repeat: matches 77..285 of consensus;
 incomplete repeat"
 334..392
 /note="MER45 repeat: matches 4..62 of consensus"
 662..959
 /note="AluSc repeat: matches 1..299 of consensus"
 1362..1544
 /note="L1PB3 repeat: matches 640..811 of consensus"
 1546..1836
 /note="AluSg repeat: matches 299..1 of consensus"
 1839..1937
 /note="L1MA4A repeat: matches 941..1046 of consensus"
 2448..2745
 /note="AluJb repeat: matches 3..301 of consensus"
 3863..4153
 /note="AluSg repeat: matches 1..293 of consensus"
 4897..4960
 /note="MIR2 repeat: matches 80..146 of consensus"
 7345..7647
 /note="AluSg repeat: matches 303..1 of consensus"
 8224..8333
 /note="MIR repeat: matches 178..66 of consensus"
 8763..8945
 /note="AluY repeat: matches 300..104 of consensus;
 incomplete repeat"
 8955..9012
 /note="AluYb8 repeat: matches 27..83 of consensus;
 incomplete repeat"
 9495..9794
 /note="AluJb repeat: matches 1..297 of consensus"
 9797..10088
 /note="AluX repeat: matches 1..301 of consensus"
 11019..11210
 /note="MIR repeat: matches 7..199 of consensus"
 11212..11507
 /note="AluX repeat: matches 1..296 of consensus"
 13312..13612
 /note="AluY repeat: matches 301..1 of consensus"
 14187..14497
 /note="AluY repeat: matches 1..301 of consensus"
 14952..15077
 /note="MIR repeat: matches 217..84 of consensus"
 15302..15475
 /note="MER5B repeat: matches 1..178 of consensus"
 15829..16294
 /note="MLTIC repeat: matches 1..466 of consensus"
 17194..17481
 /note="AluJo repeat: matches 1..302 of consensus"
 17715..18093
 /note="MSTB repeat: matches 4..361 of consensus"
 18095..18395
 /note="AluY repeat: matches 301..1 of consensus"
 18397..18461
 /note="MSTD repeat: matches 330..394 of consensus"
 20030..20334
 /note="AluSc repeat: matches 1..299 of consensus"
 23448..23750
 /note="AluX repeat: matches 1..300 of consensus"
 24157..24456
 /note="AluSg repeat: matches 1..300 of consensus"
 29337..29632
 /note="AluSg repeat: matches 296..1 of consensus"
 30246..30393
 /note="LOR1 repeat: matches 1..149 of consensus"
 30394..30691
 /note="AluX repeat: matches 1..299 of consensus"
 30692..30789

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region


```

Db 98983 AATAAATAATATTATTAT 98966

RESULT 5
AC008576/c
LOCUS
DEFINITION Homo sapiens chromosome 19 clone CTC-557J18, WORKING DRAFT
AC008576
AC008576.3 GI:6997021
VERSION
KEYWORDS
SOURCE
SOURCE
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
On Feb 18, 2000 this sequence version replaced gi:6601030.
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----Summary Statistics
Consensus quality: 285737 bases at least Q40
Consensus quality: 300872 bases at least Q30
Consensus quality: 311585 bases at least Q20
Estimated insert size: 349919; sum-of-contigs estimation
Estimated insert size: 314630; agarose-fp estimation
Quality coverage: 1.90x in Q20 bases; agarose-fp estimation
Quality coverage: 1.71x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 56 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1
* 3011: contig of 3011 bp in length
* gap of unknown length
* 3012 6593: contig of 3572 bp in length
* gap of unknown length
* 6584 9679: contig of 3096 bp in length
* gap of unknown length
* 9680 11837: contig of 2158 bp in length
* gap of unknown length
* 11838 13926: contig of 2089 bp in length
* gap of unknown length
* 13927 16002: contig of 2076 bp in length
* gap of unknown length
* 16003 18305: contig of 2303 bp in length
* gap of unknown length
* 18306 20541: contig of 2236 bp in length
* gap of unknown length
* 20542 23415: contig of 2874 bp in length
* gap of unknown length
* 23416 25390: contig of 1975 bp in length
* gap of unknown length
* 25391 27927: contig of 2537 bp in length
* gap of unknown length
* 27928 30114: contig of 2187 bp in length
* gap of unknown length
* 30115 32716: contig of 2602 bp in length
* gap of unknown length
* 32717 34880: contig of 2164 bp in length
* gap of unknown length
*
* 34881 36992: contig of 2112 bp in length
* gap of unknown length
* 36993 40002: contig of 3010 bp in length
* gap of unknown length
* 40003 42499: contig of 2497 bp in length
* gap of unknown length
* 42500 44537: contig of 2038 bp in length
* gap of unknown length
* 44538 46636: contig of 2099 bp in length
* gap of unknown length
* 46637 49202: contig of 2566 bp in length
* gap of unknown length
* 49203 51194: contig of 1992 bp in length
* gap of unknown length
* 51195 54232: contig of 3038 bp in length
* gap of unknown length
* 54233 56311: contig of 2079 bp in length
* gap of unknown length
* 56312 59177: contig of 2866 bp in length
* gap of unknown length
* 59178 61728: contig of 2551 bp in length
* gap of unknown length
* 61729 65138: contig of 3410 bp in length
* gap of unknown length
* 65139 68509: contig of 3371 bp in length
* gap of unknown length
* 68510 72140: contig of 3631 bp in length
* gap of unknown length
* 72141 74177: contig of 2037 bp in length
* gap of unknown length
* 74178 76657: contig of 2480 bp in length
* gap of unknown length
* 76658 81120: contig of 4463 bp in length
* gap of unknown length
* 81121 83944: contig of 2824 bp in length
* gap of unknown length
* 83945 88969: contig of 5025 bp in length
* gap of unknown length
* 88970 94016: contig of 5047 bp in length
* gap of unknown length
* 94017 99820: contig of 5804 bp in length
* gap of unknown length
* 99821 103739: contig of 3919 bp in length
* gap of unknown length
* 103740 110294: contig of 6555 bp in length
* gap of unknown length
* 110295 116891: contig of 6597 bp in length
* gap of unknown length
* 116892 122399: contig of 5508 bp in length
* gap of unknown length
* 122400 129404: contig of 7005 bp in length
* gap of unknown length
* 129405 136213: contig of 6809 bp in length
* gap of unknown length
* 136214 142142: contig of 5929 bp in length
* gap of unknown length
* 142143 150401: contig of 8259 bp in length
* gap of unknown length
* 150402 158946: contig of 8545 bp in length
* gap of unknown length
* 158947 167955: contig of 9009 bp in length
* gap of unknown length
* 167956 177918: contig of 9963 bp in length
* gap of unknown length
* 177919 183990: contig of 6072 bp in length
* gap of unknown length
* 183991 191939: contig of 7949 bp in length
* gap of unknown length
* 191940 201055: contig of 9116 bp in length
* gap of unknown length
* 201056 210412: contig of 9357 bp in length
* gap of unknown length

```

* 210413 219973: contig of 9561 bp in length
 * 219974 228547: gap of unknown length
 * 228548 252971: contig of 24424 bp in length
 * 252972 284497: gap of unknown length
 * 284498 316714: gap of unknown length
 * 316715 349919: contig of 33205 bp in length.

FEATURES

source
 1. 349919
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="19"
 /clone="CTC-557J18"

BASE COUNT 96739 a 81094 c 78642 g 92256 t 1188 others
 ORIGIN

Query Match 9.7%; Score 69.8; DB 54; Length 349919;
 Best Local Similarity 57.2%; Pred. No. 0.29;
 Matches 147; Conservative 0; Mismatches 107; Indels 3; Gaps 1;
 Qy 50 actaattatttgggttaataatcaattatttttaagatat---taattaagaatt 106
 Db 309365 ACAACAATCTCGTTAAATGAAATGGTTTATAATATATATATATATATATT 309306
 Qy 107 aaagatttttaaaaaatgataaattattattcatgatttttcacacattgat 166
 Db 309305 TATATATATATTTATATATATATATATATATTTTATTTATGATATATATATATATAT 309246
 Qy 167 ttgataataaataattttttttaaattcttcaaaaagtgtgcaaacacatttttagac 226
 Db 309245 TATAAT 309186
 Qy 227 atagcttggtctttcacaaagcattcatcattcaataacattaaaaattttaaac 286
 Db 309185 TTTTATATATATTTCTATATATATATATATATATTTTATAATATATATTTCTATATATATATATATAT 309126
 Qy 287 taacagtagaattctt 303
 Db 309125 ATTTTATAATATATCT 309109

RESULT

6
 AC026640 164119 bp DNA HTG 13-APR-2000
 LOCUS
 DEFINITION Homo sapiens chromosome 11 clone RP11-69K18 map 11, WORKING DRAFT
 SEQUENCE, 16 unordered pieces.
 AC026640
 AC026640.2 GI:7547222
 VERSION
 HTG; HTGS_PHASE1; HTGS_DRAFT.
 SOURCE
 human.
 ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
 Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
 Boguslavsky,L., Boukhaltier,B., Brown,A., Burkett,G.,
 Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
 Collymore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,
 Dodge,S., Domino,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D.,
 Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
 Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
 Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
 Klein,J., LaRocque,K., Lanazares,R., Landers,T., Lehoczy,J.,

Levine,R., Liu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
 McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
 Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mienga,V., Morrow,J.,
 Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
 O'Neil,D., Olivari,T.M., Oliver,J., Peterson,K., Pierre,N.,
 Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
 Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
 Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
 Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
 Vassiliev,H., Viel,R., Vo.A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
 Young,G., Zainoun,J., Zimmer,A. and Zody,M.
 Direct Submission
 Submitted (22-MAR-2000) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Apr 13, 2000 this sequence version replaced gi:7284664.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html
 ----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WITBR
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center Project name: L8749
 Center clone name: 69_K18
 ----- Summary Statistics
 Sequencing vector: M13; M7815; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.960731
 Consensus quality: 150998 bases at least Q40
 Consensus quality: 158219 bases at least Q30
 Consensus quality: 160955 bases at least Q20
 Insert size: 176000; agarose-fp
 Insert size: 162619; sum-of-contigs
 Quality coverage: 3.8 in Q20 bases; agarose-fp
 Quality coverage: 4.1 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 16 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 1608: contig of 1608 bp in length
 * 1609 1708: gap of 100 bp
 * 1709 3184: contig of 1476 bp in length
 * 3185 3284: gap of 100 bp
 * 3285 5541: contig of 2257 bp in length
 * 5542 5641: gap of 100 bp
 * 5642 7773: contig of 2132 bp in length
 * 7774 7873: gap of 100 bp
 * 7874 10690: contig of 2817 bp in length
 * 10691 10790: gap of 100 bp
 * 10791 13433: contig of 2643 bp in length
 * 13434 13533: gap of 100 bp
 * 13534 18455: contig of 2922 bp in length
 * 16456 16555: gap of 100 bp
 * 16556 21573: contig of 5018 bp in length
 * 21574 21673: gap of 100 bp
 * 21674 25705: contig of 4032 bp in length
 * 25706 25805: gap of 100 bp
 * 25806 31983: contig of 6178 bp in length
 * 31984 32083: gap of 100 bp
 * 32084 42802: contig of 10719 bp in length
 * 42803 42902: gap of 100 bp
 * 42903 54889: contig of 11987 bp in length
 * 54890 54989: gap of 100 bp
 * 54990 67682: contig of 12693 bp in length
 * 67683 67782: gap of 100 bp
 * 80005: contig of 12223 bp in length

TITLE JOURNAL COMMENT


```

* 27254 27353: gap of 100 bp
* 27354 41791: contig of 14438 bp in length
* 41792 41891: gap of 100 bp
* 41892 61037: contig of 19146 bp in length
* 61038 61137: gap of 100 bp
* 61138 88113: contig of 26976 bp in length
* 88114 88213: gap of 100 bp
* 88214 131619: contig of 43406 bp in length
* 131620 131719: gap of 100 bp
* 131720 179510: contig of 47791 bp in length.
FEATURES
  Location/Qualifiers
    1..179510
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      /clone_lib="RPC1-11 Human Male BAC"
    1..2601
      /note="assembly_fragment"
    2702..8297
      /note="assembly_fragment"
    8398..27253
      /note="assembly_fragment"
    27354..41791
      /note="assembly_fragment"
      clone_end:SP6
      vector_side:right"
    41892..61037
      /note="assembly_fragment"
    61138..88113
      /note="assembly_fragment"
    88214..131619
      /note="assembly_fragment"
      clone_end:r7
      vector_side:right"
    131720..179510
      /note="assembly_fragment"
BASE COUNT 59127 a 32374 c 31805 g 55501 t 703 others
ORIGIN

```

```

Query Match          9.6%  Score 69.2; DB 71; Length 179510;
Best Local Similarity 52.8%  Pred. No. 0.42;
Matches 149; Conservative 0; Mismatches 133; Indels 0; Gaps 0;

Qy 22 ataacatggatctccctaccacataactatttttgggttaataattcaattca 81
Db 3271 ATATAATATATATATATATATATATATATATATATATATATATATAGAATA 3212
Qy 82 ttttcaagatatattaaagattttttaaataaaagataaaattatatt 141
Db 3211 TATAATATATATATATATATATATATATATATATATATATATATATATATA 3152
Qy 142 attcatgatttttcacattgatttgataaataatttttttaattttcttaaa 201
Db 3151 TTATATATATATATATATATATATATATATATATATATATATATATATATA 3092
Qy 202 aaatgttcaagacacttattagacatgcttctgtttcacaagaacatcatt 261
Db 3091 TTATATATATCTAAATGTAATGTAATGTAATGTAATGTAATGTAATGTAAT 3032
Qy 262 taatacattaaaaatttaataactaacagtagaattctct 303
Db 3031 AATTATATATATATATATATATATATATATATATATATATATATATATATAT 2990

```

```

RESULT 8
LOCUS AC023371 48532 bp DNA HTG 14-FEB-2000
DEFINITION Homo sapiens clone RP11-21D18, LOW-PASS SEQUENCE SAMPLING.
ACCESSION AC023371
VERSION AC023371.1 GI:6970502
KEYWORDS HTG; HTGS_PHASE0.
SOURCE human.

```

ORGANISM

Homo sapiens

Eukaryota; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 48532)

Birren,B., Linton,L., Nusbaum,C. and Lander,E.

AUTHORS

Homo sapiens, clone RP11-21D18

JOURNAL

Unpublished

REFERENCE

2 (bases 1 to 48532)

Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N., Anderson,S., Baldwin,J., Baria,N., Beda,F., Boguslavsky,L., Bouckhalter,B., Brown,A., Burkett,G., Campoliano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P., Dearellano,K., Dewar,K., Dodge,S., Domino,M., Doyle,M., Fenestor,J., Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J., Gardyna,S., Glnde,S., Goyette,M., Graham,L., Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L., Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J., Landers,T., Larcocque,K., Lehoczy,J., Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Olivari,T.M., Peterson,K., Pierre,N., Pisanl,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Stojanovic,N., Severy,P., Spencer,B., Stange-Thomann,N., Theodore,J., Tirrell,A., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Vo,A., Wilson,B., Travers,M., Triglio,J., Vassiliev,H., Viel,R., Vo,A., Wyman,D., Ye,W.J., Young,G., Zalnoun,J., Zimmer,A. and Zody,M.

TITLE

Direct Submission

Submitted (14-FEB-2000)

Research, 320 Charles Street, Cambridge, MA 02141, USA

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence_submission@genome.wi.mit.edu

Project Information

Center project name: L3985

Center clone name: 21_D_18

* NOTE: This record contains 55 individual

* sequencing reads that have not been assembled into

* contigs. Runs of N are used to separate the reads

* and the order in which they appear is completely

* arbitrary. Low-pass sequence sampling is useful for

* identifying clones that may be gene-rich and allows

* overlap relationships among clones to be deduced.

* However, it should not be assumed that this clone

* will be sequenced to completion. In the event that

* the record is updated, the accession number will

* be preserved.

* 1 869: contig of 869 bp in length

* 870 gap of unknown length

* 1715: contig of 846 bp in length

* gap of unknown length

* 1716 contig of 882 bp in length

* gap of unknown length

* 2598: contig of 871 bp in length

* gap of unknown length

* 3469: contig of 901 bp in length

* gap of unknown length

* 4370: contig of 870 bp in length

* gap of unknown length

* 5240: contig of 864 bp in length

* gap of unknown length

* 6104: contig of 872 bp in length

* gap of unknown length

* 6976: contig of 896 bp in length

* gap of unknown length

BASE COUNT	404 a	83 c	33 g	455 t
ORIGIN	/gene="CAR3"			

Query Match 9.5%; Score 68.2; DB 60; Length 80920;
Best Local Similarity 53.6%; Pred. NO. 0.7;
Matches 142; Conservative 0; Mismatches 123; Indels 0; Gaps 0

[illegible]

RESULT	11
CHMPXX	121024 bp DNA circular PLN
LOCUS	16-FEB-2000
DEFINITION	Marchantia polymorpha chloroplast genome DNA.
ACCESSION	X04465 Y00686
VERSION	X04485.1 GI:11640
KEYWORDS	16S ribosomal RNA; 23S ribosomal RNA; 4.5S ribosomal RNA; 5S ribosomal RNA; ATPase; chlorophyll a apoprotein; cytochrome; cytochrome b-559; cytochrome b6; cytochrome f; cytochrome subunit IV; P-0-ATPase; F1-ATPase; initiation factor; photosynthesis; photosystem II; ribosomal protein; ribosomal protein L14; ribosomal protein L16; ribosomal protein L20; ribosomal protein L21; ribosomal protein L22; ribosomal protein L23; ribosomal protein L33; ribosomal protein S11; ribosomal protein S12; ribosomal protein S14; ribosomal protein S15; ribosomal protein S18; ribosomal protein S19; ribosomal RNA; ribulose biphosphate carboxylase; RNA polymerase; transfer RNA; transfer RNA-Ala; transfer RNA-Arg; transfer RNA-Asp; transfer RNA-Cys; transfer RNA-Gln; transfer RNA-Glu; transfer RNA-Gly; transfer RNA-His; transfer RNA-Ile; transfer RNA-Leu; transfer RNA-Lys; transfer RNA-Met; transfer RNA-Phe; transfer RNA-Pro; transfer RNA-Ser; transfer RNA-Thr; transfer RNA-Trp; transfer RNA-Tyr; transfer RNA-Val; unidentified reading frame.
SOURCE	liverwort.
ORGANISM	Chloroplast Marchantia polymorpha
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Marchantiophyta; Marchantiales; Marchantiaceae; Marchantia.
AUTHORS	1 (bases 1 to 121024)
TITLE	Yamano,Y., Ohyama,K. and Komano,T. Nucleotide sequences of chloroplast 5S ribosomal RNA from cell suspension cultures of the liverworts Marchantia polymorpha and Jungermannia subulata
JOURNAL	Nucleic Acids Res. 12 (11), 4621-4624 (1984)
MEDLINE	84247325
REFERENCE	2 (bases 1 to 121024)
AUTHORS	Umesono,K., Inokuchi,H., Ohyama,K. and Ozeki,H.
TITLE	Nucleotide sequence of Marchantia polymorpha chloroplast DNA: a region possibly encoding three tRNAs and three proteins including

JOURNAL MEDLINE REFERENCE	TITLE	
89068687 JOURNAL MEDLINE REFERENCE AUTHORS	homologue of E. coli ribosomal protein S14 Nucleic Acids Res. 12 (24), 9551-9565 (1984) 85087956	
3 (bases 1 to 121024) Fukuzawa,H., Uchida,Y., Yamano,Y., Ohshima,K. and Komano,T. Molecular cloning of promoters functional in Escherichia coli from Chloroplast DNA of a liverwort, Marchantia polymorpha Agric. Biol. Chem. 49, 2725-2731 (1985)		
4 (bases 1 to 121024) Yamano,Y., Kohchi,T., Fukuzawa,H., Ohshima,K. and Komano,T. Nucleotide sequences of chloroplast 4.5 S ribosomal RNA from a leafy liverwort, Jungermannia subulata, and thalloid liverwort, Marchantia polymorpha FEBS Lett. 185, 203-207 (1985)		
5 (bases 1 to 121024) Ohshima,K., Fukuzawa,H., Kohchi,T., Shirai,H., Sano,T., Sano,S., Umeshono,K., Shiki,Y., Takeuchi,M., Chang,Z., Aota,S., Inokuchi,H. and Ozeki,H. Chloroplast gene organization deduced from complete sequence of liverwort Marchantia polymorpha chloroplast DNA Nature 322, 572-574 (1986)		
6 (bases 1 to 121024) Fukuzawa,H., Kohchi,T., Shirai,H., Ohshima,K., Umeshono,K., Inokuchi,H. and Ozeki,H. Coding sequences for chloroplast ribosomal protein S12 from the liverwort, Marchantia polymorpha, are separated far apart on the different DNA strands FEBS Lett. 198, 11-15 (1986)		
7 (bases 1 to 121024) Fukuzawa,H., Yoshida,T., Kohchi,T., Okumura,T., Sawano,Y. and Ohshima,K. Splicing of group II introns in mRNAs coding for cytochrome b6 and subunit IV in liverwort Marchantia polymorpha chloroplast genome: Exon specifying a region coding for two genes with the spacer region FEBS Lett. 220, 61-66 (1987)		
8 (bases 1 to 121024) Kohchi,T., Ogura,Y., Umeshono,K., Yamada,Y., Komano,T., Ozeki,H. and Ohshima,K. Ordered processing and splicing in a polycistronic transcript in liverwort chloroplasts Curr. Genet. 14 (2), 147-154 (1988)		
89028845 JOURNAL MEDLINE REFERENCE AUTHORS	9 (bases 1 to 121024) Kohchi,T., Shirai,H., Fukuzawa,H., Sano,T., Komano,T., Umeshono,K., Inokuchi,H., Ozeki,H. and Ohshima,K. Structure and organization of Marchantia polymorpha chloroplast genome. IV. Inverted repeat and small single copy regions J. Mol. Biol. 203 (2), 353-372 (1988)	
89068688 JOURNAL MEDLINE REFERENCE AUTHORS	10 (bases 1 to 121024) Umeshono,K., Inokuchi,H., Shiki,Y., Takeuchi,M., Chang,Z., Fukuzawa,H., Kohchi,T., Shirai,H., Ohshima,K. and Ozeki,H. Structure and organization of Marchantia polymorpha chloroplast genome. II. Gene organization of the large single copy region from rps'12 to atpB J. Mol. Biol. 203 (2), 299-331 (1988)	
89068686 JOURNAL MEDLINE REFERENCE AUTHORS	11 (bases 1 to 121024) Ohshima,K., Fukuzawa,H., Kohchi,T., Sano,T., Sano,S., Shirai,H., Umeshono,K., Shiki,Y., Takeuchi,M., Chang,Z., Aota,S., Inokuchi,H. and Ozeki,H. Structure and organization of Marchantia polymorpha chloroplast genome. I. Cloning and gene identification J. Mol. Biol. 203 (2), 281-298 (1988)	
89068685 JOURNAL MEDLINE REFERENCE AUTHORS	12 (bases 1 to 121024) Fukuzawa,H., Kohchi,T., Sano,T., Shirai,H., Umeshono,K., Inokuchi,H., Ozeki,H. and Ohshima,K. Structure and organization of Marchantia polymorpha chloroplast genome. III. Gene organization of the large single copy region from rbcL to trnI(CAU) J. Mol. Biol. 203 (2), 333-351 (1988)	
89068684 JOURNAL MEDLINE REFERENCE AUTHORS	13 (bases 1 to 121024) Shimada,H. and Sugiyama,M. Fine structural features of the chloroplast genome: comparison of the sequenced chloroplast genomes Nucleic Acids Res. 19 (5), 983-995 (1991)	
91212240 JOURNAL MEDLINE REFERENCE AUTHORS	14 (bases 1 to 121024) Ohyama,K. Direct Submission Submitted (17-OCT-1986) K. Ohyama, Research Centre for Cell and Tissue Culture, Faculty of Agriculture, Kyoto University, Kyoto 606, Japan GENERAL DESCRIPTION OF SEQUENCE gene name product name Ribosomal RNA genes 4.5S rRNA 5S rRNA RNA polymerase genes alpha subunit subunit subunit subunit Ribosomal protein genes and related genes 50S subunit rpl2: homologous to E. coli L2 L14 rpl16: homologous to E. coli L1 rpl20: homologous to E. coli L20 rpl21: homologous to E. coli L21 rpl22: homologous to E. coli L22 rpl23: homologous to E. coli L23 rpl33: homologous to E. coli 30S subunit rps2: homologous to E. coli S2 rps3: homologous to E. coli S3 rps4: homologous to E. coli S4 rps7: homologous to E. coli S7 rps8: homologous to E. coli S8 S11 rps11: homologous to E. coli S11 rps12: homologous to E. coli S12 rps14: homologous to E. coli S14 rps15: homologous to E. coli S15 rps18: homologous to E. coli S18 rps19: homologous to E. coli other genes homologous to E. coli Genes for photosynthesis rbcL: large subunit of Rubisco psaA: photosystem I P700 chlorophyll (an apoprotein) psaB: same as above psaD: photosystem II 32kd protein psbA: photosystem II p680 chlorophyll (an apoprotein) psbC: same as above psbD: photosystem II D2 protein psbE: cytochrome b559 psbF: same as above psbG: photosystem II G protein atpA: ATPase F1 subunit alpha atpB: ATPase F1 subunit beta atpE: ATPase F0 subunit I atpF: ATPase F0 subunit II atpH: ATPase F0 subunit III atpI: ATPase F0 subunit IV petA: cytochrome f petB: cytochrome b6 petD: subunit 4 of cytochrome b6/f complex	


```
/evidence-not_experimental
/product="dull08D11.1 (novel protein similar to C. elegans
T22C1.7 (Tr:Q22667) (part of protein dull183I21.2))"
/protein_id="CAB81659.1"
/db_xref="GI:7263911"
/translation="YYQQTNGMHRGCVGVSYPVGMVVRPLRTSLSLRSEHS
NCTVADSPASDPSPALPSPAIPIRGGFALSLANAAARAPKGGIFQFGALLGK
LRTAESVSGSRVSFLKSLSSGASDAASTASLGEAAEGADEAAPPFEADIDATT
TTEYGNWKNRSGFGVYSERSGLRYEGEDLNRHGCGCTTLPDGHRECKYRHHV
LVKDTKRLMLQKSKNKVKQKVEHVSVEGAQRAAAIARAKAEIATSTSHAKAKAEAEQ
AALAAQESNTARTLARELPDYCP"
repeat_region 3588..3794
/feature="3 copies 69 mer 72% conserved"
repeat_region 4958..5135
/feature="MIR repeat: matches 69..252 of consensus"
repeat_region 5428..5566
/feature="MIR repeat: matches 113..254 of consensus"
repeat_region 5733..5845
/feature="MIR repeat: matches 25..136 of consensus"
repeat_region 5846..5921
/feature="MIR repeat: matches 63..139 of consensus"
repeat_region 5977..6392
/feature="L1R16A repeat: matches 16..450 of consensus"
misc_feature complement(5998..6470)
/feature="match: GSS: Em:AQ812840"
repeat_region 6469..6548
/feature="MIR repeat: matches 63..142 of consensus"
repeat_region 6709..7083
/feature="L1R1A1 repeat: matches 3..394 of consensus"
repeat_region 7106..7166
/feature="MER94 repeat: matches 4..64 of consensus"
repeat_region 7179..7493
/feature="L1MD3 repeat: matches 7431..7739 of consensus"
repeat_region 7494..7789
/feature="AlusX repeat: matches 1..296 of consensus"
repeat_region 7790..8083
/feature="L1MD3 repeat: matches 7034..7431 of consensus"
repeat_region 8189..8388
/feature="MIR repeat: matches 8..233 of consensus"
repeat_region 9046..9363
/feature="MER33 repeat: matches 1..324 of consensus"
repeat_region 10136..10308
/feature="MIR repeat: matches 47..226 of consensus"
repeat_region 10950..11011
/feature="MIR repeat: matches 192..250 of consensus"
repeat_region 11012..11147
/feature="BC200 repeat: matches 3..136 of consensus"
repeat_region 11153..11335
/feature="MIR repeat: matches 11..201 of consensus"
repeat_region 11353..11800
/feature="L2 repeat: matches 2234..2705 of consensus"
repeat_region 12005..12308
/feature="AlusX repeat: matches 1..304 of consensus"
repeat_region 12374..12467
/feature="L1MCS repeat: matches 7833..7925 of consensus"
repeat_region 12487..12799
/feature="AluJB repeat: matches 1..309 of consensus"
repeat_region 12830..13292
/feature="L1M4 repeat: matches 4351..4825 of consensus"
repeat_region 13358..13529
/feature="L1MCS repeat: matches 7249..7419 of consensus"
repeat_region 13552..13845
/feature="AlusX repeat: matches 2..309 of consensus"
repeat_region 14067..14098
/feature="16 copies 2 mer ac 84% conserved"
repeat_region 14219..14348
/feature="L2 repeat: matches 1995..2139 of consensus"
repeat_region 14432..14761
/feature="AluY repeat: matches 1..310 of consensus"
repeat_region 15256..15341
/feature="MIR repeat: matches 156..251 of consensus"
repeat_region 15342..15662
/feature="AlusX repeat: matches 1..311 of consensus"
```

```
repeat_region 15663..15739
/feature="MIR repeat: matches 40..156 of consensus"
repeat_region 15903..16262
/feature="L1ME3A repeat: matches 5678..6018 of consensus"
misc_feature complement(16083..16468)
/feature="match: GSS: Em:AQ262524"
repeat_region 17211..17513
/feature="AluB8 repeat: matches 1..310 of consensus"
misc_feature complement(17463..17840)
/feature="match: GSS: Em:AQ776838"
repeat_region 17657..17742
/feature="MIR repeat: matches 59..144 of consensus"
repeat_region 17919..18220
/feature="AlusX repeat: matches 1..303 of consensus"
repeat_region 18485..18670
/feature="MIR repeat: matches 46..228 of consensus"
repeat_region 18705..18811
/feature="MIR repeat: matches 123..234 of consensus"
misc_feature complement(18708..19271)
/feature="match: GSS: Em:AQ760123"
misc_feature complement(18812..19279)
/feature="match: GSS: Em:AQ777683"
misc_feature complement(18825..19251)
/feature="match: GSS: Em:B90590"
repeat_region 18964..19192
/feature="MIR repeat: matches 6..247 of consensus"
repeat_region 19623..19802
/feature="L2 repeat: matches 1997..2185 of consensus"
repeat_region 19812..19999
/feature="L1R16C repeat: matches 202..384 of consensus"
repeat_region 20267..20585
/feature="L2 repeat: matches 2206..2545 of consensus"
repeat_region 20733..21144
/feature="L2 repeat: matches 2264..2706 of consensus"
repeat_region 21176..21340
/feature="MER33 repeat: matches 10..205 of consensus"
repeat_region 21735..22030
/feature="AluJ repeat: matches 1..296 of consensus"
repeat_region 22059..22239
/feature="MLT1J repeat: matches 250..439 of consensus"
repeat_region 22345..22416
/feature="MLT1J repeat: matches 1..72 of consensus"
repeat_region 22522..22556
/feature="MLT1J repeat: matches 1..72 of consensus"

Query Match 9.3% Score 66.8; DB 11; Length 152209;
Best Local Similarity 52.5%; Pred. No. 0.88;
Matches 146; Conservative 0; Mismatches 132; Indels 0; Gaps 0;

QY 27 atggatatctccctaccatcatactaatattttgggttaaatatcattattttt 86
Db 29312 ATATATATTTAATAAATAATATATATTTATATATATATATATATATAT 29371

QY 87 aagatattaataaagaatttttaaaaaaaatgataaaattatatttca 146
Db 29372 TAAATATTTATATTTAATAAATAATTTATTTATATATTTAATAAATAAT 29431

QY 147 tgattttcatacattgattttgataaataatatttttttaattttcttaaaatg 206
Db 29432 TATATTTAATAAATAATATATTTAATAAATAATATATTTATATATTTAATAAATAT 29491

QY 207 ttcaagacactatttagacatagcttctgtgtttacaagaatcattcatc 266
Db 29492 ATATTTATATATAAATAATATATTTATATATATTTAATAAATAATATATTTATATATAT 29551

QY 267 cattaaaaaatttaataactaacagtagaattctttt 304
Db 29552 TAAATAATATATTTATATATATATATATATATATATATATATATATATAT 29589
```

RESULT 15

MTSCAJ23

LOCUS

DEFINITION

MTSCAJ23 1867 bp DNA circular PLN 11-MAR-1998
Saccharomyces cerevisiae mitochondrial rRNA-Tyr, tRNA-Asn, &

Search completed: September 2, 2000, 03:05:39
Job time: 12803 sec

[illegible]